

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:06:59 ; Search time 24.81 Seconds
(without alignments)
1385.329 Million cell updates/sec

Title: US-09-426-072-2
Perfect score: 2501
Sequence: 1 MKSFALTGLGLSLGLAS.....GTTLRDAPIALKTPHTFSV 464

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues 522463

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2501	100.0	464	22 AAB04146	Phospholipase B of
2	126.5	5.1	483	11 AAR07445	Secretory signal P
3	111.5	4.5	1041	18 AAW11866	ATH1 gene product,
4	109.5	4.4	579	14 AAR38697	PSI875 swine parvo
5	109.5	4.4	925	21 AAB23735	Agarase 1-7 protei
6	109.5	4.4	1041	20 AAY16788	S. cerevisiae acid
7	108.5	4.3	626	21 AAG30820	Arabidopsis thalia
8	107	4.3	532	21 AAY58162	Adeno associated v
9	107	4.3	588	21 AAY58161	Adeno associated v
10	107	4.3	724	21 AAY58160	Adeno associated v
11	107	4.3	1257	13 AAR29028	Bacillus thuringie

12	107	4.3	1257	13 AAR28811	BT toxin 33F2. Ba
13	107	4.3	1257	13 AAR29518	BT toxin 33F2. Ba
14	107	4.3	1257	14 AAR44203	Bacillus thuringie
15	107	4.3	1257	18 AAW13886	33F2 toxin. Bacil
16	107	4.3	1257	21 AAB13893	Bacillus thuringie
17	107	4.3	1257	22 AAB59883	33F2 protein. Bac
18	106	4.2	1257	13 AAR20068	B thuringiensis to
19	105.5	4.2	962	21 AAB18211	Plasmodium falcipa
20	105.5	4.2	1041	20 AAY29830	Cyathosin protein
21	105.5	4.2	1122	16 AAR64927	Arg-gingipain-1.
22	104.5	4.2	737	16 AAR70186	Arg-gingipain high
23	104.5	4.2	737	18 AAW34846	Arg-gingipain-2 am
24	104.5	4.2	737	21 AAY67395	porphyromonas ging
25	104.5	4.2	991	16 AAR77313	P. gingivalis haem
26	104.5	4.2	1687	17 AAR96033	Haemagglutinin pro
27	104.5	4.2	1687	19 AAW69495	Arg-gingipain-2 pr
28	104.5	4.2	1704	16 AAR70188	Arg-gingipain high
29	104.5	4.2	1704	18 AAW34843	Arg-gingipain-2 am
30	104.5	4.2	1704	21 AAY67396	Neisseria meningit
31	104	4.2	783	21 AAY75533	Neisseria meningit
32	104	4.2	783	21 AAY75534	Eucalyptus grandis
33	103.5	4.1	414	21 AAB33292	Swine parvovirus B
34	103.5	4.1	579	14 AAR38702	Swine parvovirus B
35	103.5	4.1	579	17 AAR99721	pig parvo virus B
36	103.5	4.1	598	8 AAP70500	Arabidopsis thalia
37	103.5	4.1	1040	21 AAG30935	Arabidopsis thalia
38	103.5	4.1	1052	21 AAG30934	Arabidopsis thalia
39	103.5	4.1	1181	21 AAG30933	Rat von Ebner's gl
40	103.5	4.1	1290	18 AAW07609	Bacillus sp. alkal
41	103	4.1	636	20 AAW89548	Sequence of a porc
42	103	4.1	686	5 AAP40068	Alkali-protease ya
43	102	4.1	433	13 AAR26274	Modified Bacillus
44	102	4.1	433	19 AAW61495	Bacillus sp. Lion
45	102	4.1	433	20 AAW95698	

ALIGNMENTS

RESULT 1
AAB04146
ID AAB04146 standard; Protein; 464 AA.
XX
AC AAB04146;
XX
XX 11-APR-2001 (first entry)
XX
XX Phospholipase B of Aspergillus oryzae.
DE
XX Phospholipase B: hydrolysis; lysophospholipid;
KW Phospholipid; lecithin; lysolecithin; degumming; starch;
KW wheat starch; oil; dough; baking; strength; elasticity;
KW stability; extensibility.
XX
XX Aspergillus oryzae.
OS
XX
XX US6146869-A.
PN
XX
XX 14-NOV-2000.
PD
XX
XX 21-OCT-1999; 99US-0426072.
PF
XX
XX 21-OCT-1999; 99US-0426072.
PR
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
PA
XX
XX Harris P, Brown KM;
PI
XX
XX WPI; 2001-049057/06.
DR
XX
XX N-PSDB; AAA54535.
PT
XX
XX Novel polypeptide having phospholipase B activity useful for
PT hydrolysing fatty acyl groups of phospholipid, preparing dough or baked

PT product and reducing phospholipid content in an edible oil
 XX Claim 1; Fig 1a-b; 24pp; English.
 PS
 CC Phospholipase B is useful for hydrolysing the fatty acyl group(s) of
 CC phospholipids or lysophospholipids, such as lecithin or lysolecithin,
 CC degumming an aqueous carbohydrate solution or slurry to improve its
 CC filterability, particularly, a starch hydrolysate or wheat starch
 CC hydrolysate which is difficult to filter and to reduce the
 CC phospholipid content in an edible oil by treating the oil with
 CC phospholipase B to hydrolyse a major portion of the phospholipid
 CC and separating an aqueous phase containing the hydrolysed
 CC phospholipid from the oil. Phospholipase B is also useful for
 CC preparing a dough or a baked product having improved property such
 CC as increased strength, elasticity, stability and extensibility.
 XX
 SQ Sequence 464 AA;

Query Match 100.0%; Score 2501; DB 22; Length 464;
 Best Local Similarity 100.0%; Pred. No. 1e-215;
 Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSTALLTGGLASGASPVTSVTSREAPFGYKPGKESIKENLKDKNVNIWILE 60
 DB 1 mkstalltgglasgiaspvtsevtvseapfykpgskesenlkdkvenivwile 60
 QY 61 NRSFDNILGGVRRQGLDNPINNPGFCNKNASDPSSGKYCTQAKDYDSVFNDPDHSTGN 120
 DB 61 nrfdnlgvrrqgldnpinnpgfcnyknasdpssgkyctqakdydsvfndpdhstgn 120
 QY 121 NLEPYGYTPNNKATASGVADQSGFLNAQLNDYKLAPEEATRVNGYITEEVTPLV 180
 DB 121 nlepygytpnngaasgvvadqsgflnaqlndykpklapeeatrvngyiteevtplv 180
 QY 181 DLVDEFTTNSFCVPGTNPRLCALAGTAAGHGKNDDEFLNYGISKSIPEAAKEG 240
 DB 181 dlvdetttnsfscvpgtpnprlcalagtaaghgknddflnygisksipeaanekeg 240
 QY 241 VSWLNDGTNGEFPDSLFTFYVNTSRNVVPEFODAYLGVLPKESINSCCGTN 300
 DB 241 vswlndgtnggefdsllftfyvntsrnvvpvpefodaylgyvpkfsynpsccgtn 300
 QY 301 TNSMHPGTNVSFGEVFKQYDAIRGQPQWDTLLFTYDETGGFYDHVPPPLAVRPDNL 360
 DB 301 tnsmhptgnvsfgevfkqydairgqpqwdtllfitydetggfydhvppplavrpdl 360
 QY 361 TYTETAKNGKYTLHFDRLGGRMPTWVISPXSKGYEIOXGTDPTVTKGPAPYSATSVLKT 420
 DB 361 tytetakngkytlhforlgrmptwvispyskgyeioxgtdptvtkgpapysatvikt 420
 QY 421 LGYLWDIEDFTPRVAHSPFDHLIGTLREDAPTALKTPTFSV 464
 DB 421 lgyldiedftprvahspfdhlgtlredaptalktptftsv 464

RESULT 2
 AAR07445
 ID AAR07445 standard; protein; 483 AA.
 XX
 AC AAR07445;

DT 01-FEB-1991 (first entry)

DE Secretory signal peptide and protein with repressible acid
 DE phosphatase activity.

KW Acid phosphatase; yeast.

XX Saccharomyces cerevisiae.

OS
 FH Key Location/Qualifiers
 FT Peptide 1..21

FT
 FT Protein /label=Secretory signal peptide
 FT 22..483
 XX /label=Protein has repressible acid phosphatase activity
 XX JP02234681-A.
 XX 17-SEP-1990.
 XX 09-MAR-1989; 89JP-0057234.
 XX 09-MAR-1989; 89JP-0057234.
 XX (IKEM) IKEDA MORANDO KK.
 XX WPI; 1990-325618/43.
 XX N-PSDB; AAQ06266.
 XX DNA contg. base sequences with promoter activity - used for
 XX expression and secretory prodn. of pharmacologically important
 XX protein gene
 XX Claim 1; Page 553; lpp; Japanese.
 XX Protein has repressible acid phosphatase activity, and carries a
 XX secretory signal region allowing it to be produced in high yield into
 XX the extracellular environment.
 SQ Sequence 483 AA;

Query Match 5.1%; Score 126.5; DB 11; Length 483;
 Best Local Similarity 23.0%; Pred. No. 0.011;
 Matches 67; Conservative 47; Mismatches 112; Indels 65; Gaps 14;

QY 134 AIASGKVADQ-SGLNAQLN-----DYPKLAPEEATRVNGYITEEVTPLVDL 182
 DB 48 aaasavahtrtdskvaavrfyflentdydkavede-----slawl 92
 QY 183 VDEFTTNSFCVPGTNPRLCALAGTAAGHGKNDDEFLNYGISKSIPEAAKEGYS 242
 DB 93 akqgisltnyws-lthpsepnylasvagdy--faldddrfismpsvsnvldldtknls 149
 QY 243 W-----LNYDGTWG-----EFEPDSLFTFYVNTSR-SNVVPEVNFQD 280
 DB 150 waeyqehipygfgfngynsqetfandyvrkhnplllfdnvisdkrlaniksfeinfnd 209
 QY 281 AYLGVLKPEFINSFCGNTNSMHPGTNVSFGEVFKQYDAIRGQPQWDK-TLLFTY 339
 DB 210 lknktlpgyafitpnm-----tndghds-nikvagdwskfllqpllsdddyfmdktlvtlf 264
 QY 340 D--ETGGFYDHVPPPL--AVRPDNLTYTETAKNGKYTLHFDRLGGRMPTW 386
 DB 265 denetygiknkvfslilggvipdhkgt----sddtfydhysqlatveanw 311

RESULT 3

AAW11866
 ID AAW11866 standard; Protein; 1041 AA.

XX
 AC AAW11866;

DT 18-APR-1997 (first entry)

DE ATH1 gene product, Athlp.

KW Yeast ATH1; Athlp; vacuolar acid trehalase; hydrolysis; trehalose;
 KW baking; frozen dough; dehydrated yeast; brewing; ethanol fuel;
 KW molasses; corn syrup; freezing; wine; fermentation.

XX Saccharomyces cerevisiae.

OS
 FH Key Location/Qualifiers
 FT Peptide US5587290-A.

PD 24-DEC-1996.
 PF 26-JUN-1995; 95US-0494714.
 PR 26-JUN-1995; 95US-0494714.
 XX (REGC) UNIV CALIFORNIA.
 XX Destruelle M, Holzer H, Klionsky D;
 DR WPI: 1997-064793/06.
 DR N-PSDB; AA761372.
 XX Mutant yeast deficient in prodn. of vacuolar acid trehalase - has
 PT improved stress tolerance, esp. for dehydration, freezing and
 PT alcohol concn., useful in baking, fermentation, etc.
 XX
 PS Disclosure; Column 17-24; 17pp; English.
 CC This sequence is encoded by the yeast ATH1 coding sequence and represents
 CC Athlp which is a vacuolar acid trehalase. Yeast mutants which are
 CC deficient in the production of functional Athlp have reduced ability
 CC to hydrolyse trehalose. These mutants are useful in baking (partic. in
 CC frozen dough and dehydrated yeast products), in brewing, as a source of
 CC trehalose (used as a protectant in foods and pharmaceuticals) and for
 CC production of ethanol fuel from molasses or corn syrup. The yeast
 CC mutants have improved tolerance of dehydration and freezing. They can
 CC tolerate higher levels of ethanol and can grow at high cell densities
 CC over a range of fermentable sugar concentrations, and so can produce
 CC drier wines and complete fermentation more quickly.
 XX
 SQ Sequence 1041 AA;

Query Match 4.5%; Score 111.5; DB 18; Length 1041;
 Best Local Similarity 19.7%; Pred. NO. 0.84;
 Matches 89; Conservative 64; Mismatches 140; Indels 159; Gaps 24;

QY 22 VTSEY---TSVREAPFGYKPGSKESTENLKRVENIVWILENRS-----FDNLGG 70
 DB 446 vdeyhlnvdamasfilylnghedld--eylryttwplknaaqffaykykynslgl 503
 QY 71 VRROGLDNP-----INNGPCNYKNASDPSSGKYCTQAKDYDSVFNPDHSHVGTGNNLEF 124
 DB 504 yetnltddpdefanhnngaf---tnagiktlkwtadighnlghevdpkwselskdl-- 558
 QY 125 YGVTTPNNGAIAAGKVVADQSGFLNAQLNDYKLAPEATRQV--MGYVTEEVPTLVDL 182
 DB 559 ---yipr---ssnitleysg-----mssveikqadvtlmvyipgyindesi--lna 604
 QY 183 VDEFTTFNSWFCV-PGPTNPRLCALAGTAAGH-----K 217
 DB 605 ikdlyyserqasgamtvyfvaaag-llnhgsasqsylykswlpylrapfaqfseq 663
 QY 218 NDDDFLNYGSSKSI-EEAANEK-----GVSWLNYDGTNGEPEPDSLFYTYVNOTSRS 269
 DB 664 sdnflngltqapaflltagtggflsgilfgitgiry---syevdpdt---kkinrlirf 717
 QY 270 NVYPVENFODAYLGVLP-----KFSYINPSCCGTNTNSMHPGTGNVSYGEVFKQIYDA 323
 DB 718 n-pie-----lplpggiairnfkymnp----- 739
 QY 324 IROGPQWDXTLFLFYDETFGFDH-----VPPPLAVRPDLNLTVTETAKNGOKYT 373
 DB 740 -----vldliidhngtviwhksgdvpvhlkiprslrhqdqinfngsenerkpn 789
 QY 374 LH---FDRLGGRMTWVISPYSKGYIEQYGT 402
 DB 790 lerrdvrvgdpm-----rmdrygt 809

RESULT 4
 AAR38697

ID AAR38697 standard; Protein; 579 AA.
 XX
 AC AAR38697;
 XX
 DT 25-NOV-1993 (first entry)
 XX
 DE PSY875 swine parvovirus B gene product.
 XX
 KW Attenuated; vaccine; herpes virus; non-primate; live; safer; IBR;
 KW infectious bovine rhinotracheitis; MDV; Marek's disease virus; fowl;
 KW pseudo-rabies; swine.
 XX
 OS Swine parvovirus.
 XX
 PN US5223424-A.
 XX
 PD 29-JUN-1993.
 XX
 PF 27-JUL-1988; 88US-0225032.
 XX
 PR 06-SEP-1985; 85US-0773430.
 PR 27-JAN-1986; 86US-0823102.
 PR 17-JUL-1986; 86US-0887140.
 PR 02-SEP-1986; 86US-0902887.
 PR 20-NOV-1986; 86US-0933107.
 PR 27-JUL-1987; 87US-0078519.
 PR 27-JUL-1988; 88US-0225032.
 XX
 PA (PRUT-) PRUTECH RES & DEV.
 XX
 PI Chiang CH, Cochran MD, Macdonald RD;
 XX
 DR WPI: 1993-219585/27.
 DR N-PSDB; AA042754.
 XX
 PT Recombinant fusion proteins for vaccine - comprises antigenic
 PT sequences fused to viral sequences e.g. pseudo-rabies virus, used
 PT as vaccines
 XX
 PS Disclosure; Fig 11; 127pp; English.
 XX
 CC The sequence is that encoded by the swine parvovirus B gene
 CC sequence from PSY875 which may be used in the prodn. of attenuated
 CC non-primate herpes viruses. These can be used as live vaccines and
 CC provide a safer vaccine than currently available for e.g. pseudorabies
 CC virus of swine, infectious bovine rhinotracheitis (IBR) virus or Marek's
 CC disease of fowl.
 XX
 SQ Sequence 579 AA;

Query Match 4.4%; Score 109.5; DB 14; Length 579;
 Best Local Similarity 21.4%; Pred. NO. 0.51;
 Matches 86; Conservative 44; Mismatches 151; Indels 121; Gaps 22;

QY 110 FNPDHSHVGTGNNL-----EFY-----GTYTP-----NNGAIAAGKVVADQ 145
 DB 121 fnpadwqlisnmnteinlvsefeifnvkkitatesatpsklynnnditaslmvaldn 180
 QY 146 GFLNAQLNDYKLAPEATEATRQVMGYTEBEVPTLVDLVEFTTFNSWFCVPGTNPRL 205
 DB 181 ntl-----pytpaaprset---lgfy--pwlp-----kptgyryylscirnlpp--- 221
 QY 206 CALAGTAAGHGNDDDFLNYGSSKSI-EEAANEKGVSWLNYDG---TNG--EFEPDSLFF 260
 DB 222 -----tytqsgsqidsqtgihsdmftyienavpihlrtgdefstgiyhfctkplk 276
 QY 261 TYVNOTSRSNVYPVENFFQ-----DAYLGVLPKFSYINPSCCGTNT-NSMHPGTGNVSYGE 314
 DB 277 thswqtntslglppkvltptegdqhgptlp-----gantrkyghqhtnnsyte 326
 QY 315 VFVQIYDAIROGP-QWOKTLLFLFYDETFGFDHVPPLAVRPDLNLTVTETAKNGOKYT 373

XX PS Examples; Page 16-23; 57pp; English.

XX CC The invention relates to an acid trehalase (ATH1) gene-disrupted haploid

CC yeast produced by gene manipulation, where the diploid is a practical

CC baker's yeast. The new yeast strain is useful in dough which, upon

CC subsequent fermentation and baking after a period of freezing, produces

CC a high quality bread or sweetened bun. The new yeasts have a reduced

CC ability of trehalose degradation during fermentation and thus an

CC increased intracellular osmotic pressure, providing resistance to long

CC -term freeze-storage when used in low and high-sugar dough. The dough

CC makes better bread than dough using prior art baker's yeast. The new

CC strain and ATH1 non-disrupted gene strains were used to prepare dough,

CC and the new strain showed resistance to freezing. The present sequence

CC represents an ATH1 gene product.

XX SQ Sequence 1041 AA;

Query Match 4.4%; Score 109.5; DB 20; Length 1041;

Best Local Similarity 19.5%; Pred. No. 1.3;

Matches 88; Conservative 65; Mismatches 140; Indels 159; Gaps 24;

QY 22 VTSEY---TSVREAPPFGYKPGSKESIENLKDKNVENIWLLENRS-----FDNITLGG 70

DB 446 vdeyhlnvdvamasfsiylnghgidd-eylryttwpliknaaagftaykynsslgl 503

QY 71 VRRQGLDNP-----INNGPFCNYKNADSPSGKYCTQAKDYDSVFNPDHSHVGTGNLLEF 124

DB 504 yetynltdpdefahinnaf---tnagiktlkwatdighlgevdvdpkwseiskdi-- 558

QY 125 YGYTPNNGATASGVKQVADQSGFLNAQLNDYPKLAPAEATROV--MGYYTEEEVPTLVLDL 182

DB 559 ---yipr-----sssnitleysg-----mnsseviekqadvltmvyplgyindesi--lnna 604

QY 183 VDEFTFNFSWSCV-PGFTNPNRLCALAGTAAGHG-----K 217

DB 605 ikdlytyserqasgpantypvfvaaag-llnngsssqsylykswlpyrapfaqfseq 663

QY 218 NDDDFLYNGISSKST-----FEAANEKGVSWLNDYGTNGEFEPDSLFFTYVYNTQTSRS 269

DB 664 sddnftlgtapafitanggflqsnlfgitgiry---syevdpdt---kkinrlrrf 717

QY 270 NVVPVENFQDAYLGLVP-----KFSYINPSCCCTNTNSMHPGTGNVSYGVEFVKQIYDA 323

DB 718 n--pie-----lplpgglairnfkymn-----qvld- 742

QY 324 IROGPQWKTLLFIYDGTGFDH-----VPPPLAVRPDNLTYTETAKNGOKYT 373

DB 743 -----liidhngtikhsgdvpbhikpnrslihdqdnfngysenerkpn 789

QY 374 LH---FDRGGRRPTWVISPYSKGVIEQYT 402

DB 790 lerrdvrdvdpn-----rmdrygt 809

RESULT 7

AAAG30820

ID AAAG30820 standard; Protein; 626 AA.

XX AC AAAG30820;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 36914.

XX KW protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136392.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137222.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139859.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

```
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149820.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.3%; Score 108.5; DB 21; Length 626;
Best Local Similarity 21.3%; Pred. No. 0.7;
Matches 99; Conservative 55; Mismatches 183; Indels 127; Gaps 23;

QY 38 PGSKESIEMLKDVENIVWLILENRSFONILGVRROGLDNPINNGPCFNK---NASDP 94
Db 102 pdlgaslekhddvk---fsyvdqknftn-ygnaraggads-----fknyskdgnvtd 151
QY 95 SSCKTCYQAKDYDSVFN--DPDHSVTGNLFEYGYTTPNNGAIAAGKVVADQSGFLNAQL 152
Db 152 sfrrysrnaaghdhdkftvygensnvveegfusygtf-----gtggagdftnygnvnnpts 207
QY 153 NDYPKLAPPEATROVMGYTTEEEVPTLVDLVDFTTFNSWFCVPGTPNPNRLCALAGTA 212
Db 208 rftaysdggngrsqtfktythe-----anagng 235
QY 213 ---AGHGN---DDDFLNGISSKSIPEAANEKGVSWLNDGTNGEPEPDSLEFFTVNQ 265
Db 236 qsfstygkngngvneftsylvssnvi-----gsfisy-gesgnaand-----tftsy 283
QY 266 TSSNNVVFVNEFFQDAYLG--VLPKF-SYINPSCCGTNTNSMHPGTG-----VSYGEV 315
Db 284 gsdgn-vpqnfnnygasnaavdtfanyrdkanvgddsfssyakdskseknfvnygs 342
QY 316 F-VKQIYDAIROQPDWKTLFIYDETGCGFDYHVPPLAVRPNLTYTTAKNGOKYT 373
Db 343 fnpgeetfgygkaegsk-lsfktytpnstfkdyakkgvafakynvsttattvvgdgt 401
QY 374 LH-----EDRLGGR--MPTWIS--PY--SKKGIEQYG- 401
Db 402 vnkwiepgkffresslkegtvimpdkdkmpkrslprslitkpfstsklgeikrih 461
QY 402 -----TDPVTCGPAPYSATSVLTKLGLWDIEDFTPRV 434
```

Db 462 avenstmggiitdavtecerppsygetkrcvgsaedmidfatsv 505
 RESULT 8
 AAY58162
 ID AAY58162 standard; Protein: 532 AA.
 AC AAY58162;
 XX 07-MAR-2000 (first entry)
 DT Adeno associated virus AAV5 capsid protein VP3.
 DE Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;
 XX promoter; Rep protein; capsid protein; regulation; transcription;
 KW replication; chromosomal integration; tissue tropism; cellular receptor;
 KW gene therapy; neutralising antibody; erythroid progenitor cell;
 KW transduction; cancer; genetic disease; VP3.
 XX Adeno associated virus type 5.
 OS WO9961601-A2.
 XX 02-DEC-1999.
 XX 28-MAY-1999; 99WO-US11958.
 XX 28-MAY-1998; 98US-0087029.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chlorini JA, Kotin RM;
 PI WPT; 2000-062707/05.
 DR N-PSDB; AA249214.
 XX Adeno-associated virus 5 based vectors and particles, useful for gene
 PT therapy .
 PS Claim 27; Page 80-82; 91pp; English.
 XX This sequence represents the VP3 capsid protein of
 CC adeno associated virus type 5 (AAV5). The invention relates
 CC to vectors comprising a pair of AAV5 inverted terminal repeats
 CC (ITRs) with a promoter between the ITRs. The vector may comprise
 CC the viral genome, or subregions thereof, including sequences
 CC encoding Rep proteins and capsid proteins, and is encapsidated
 CC in an AAV5 particle. The non-structural Rep proteins Rep40
 CC (AAY58163), Rep52 (AAY58168), Rep68 (AAY58164) and Rep78 (AAY58159) are
 CC involved in regulation of replication and transcription, in addition to
 CC the production of progeny genomes. Rep68 and Rep78 are also associated
 CC with the stable integration of the viral genome into human chromosomes.
 CC The three types of capsid protein VP1 (AAY58160), VP2 (AAY58161) and VP3
 CC (AAY58162) assemble to form an icosahedral capsid, and differ from each
 CC other by the use of alternative splicing and an unusual translation
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are
 CC likely to utilise distinct cellular receptors and are serologically
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for
 CC transduction of a patient who already possess neutralising antibodies
 CC either as a result of natural immunological defence or from prior
 CC exposure to AAV2 vectors. The vectors may be useful for transducing
 CC erythroid progenitor cells or cells lacking heparin sulphate
 CC proteoglycans, which is very inefficient with AAV2-based vectors.
 CC The vectors may also be useful for transducing cells with a nucleic
 CC acid of interest in order to produce cell lines that could be used to
 CC screen for agents that interact with the gene product of the nucleic
 CC acid of interest. In addition to transduction of other cell types,
 CC transduction of erythroid cells would be useful for the treatment of
 CC cancer and genetic diseases which can be corrected by bone marrow
 CC transplants using matched donors.
 XX

SQ Sequence 532 AA;
 Query Match 4.3%; Score 107; DB 21; Length 532;
 Best Local Similarity 19.1%; Pred. No. 0.74;
 Matches 66; Conservative 35; Mismatches 113; Indels 132; Gaps 16;
 Qy 118 TGNLE-----FYGYTFNNGAIASGVYADQSGFLNAQLNDYKPLAPEATQVM 168
 Db 206 tgnlfeytfvphssflapnlfkianplvdqlylrfvstnn---tggvqfknla 262
 Qy 169 GYITEEVPTLVLDLVEFTFNSWFCVPGPTNPNRLCALACTAAGHGKNDDDLFLYNGIS 228
 Db 263 gya-----ntyknwf---pgpmgrtg-----gwnlgsqvnz-----a 292
 Qy 229 SKSFEANEKGVSWLNDYDCTNGEPEPDSLFTFVNTSRSNVVPVENFF----QDAYLG 284
 Db 293 svsafatnr-----mlelasyqvppqngmt---nnlgsntyalentmifnsqpanpg 345
 Qy 285 VLPKFSYINPSCCGTNTNSMHPGTGNVSY-----GEV 315
 Db 346 tta--tyleghmlitsetqpvrnvaynvvggmatnngssttapatgtynlqelvpqsv 403
 Qy 316 FVKQIYDAIROGPQWDKTLFTYDETFGFY-----DHVPPPLAVR-----PDNL 360
 Db 404 wmer--dvylgplwak-----ipetgahfhpspamggfghkppmmlikntpvgg 455
 Qy 361 TYTETAKNGOKYTLHFDRLGGRMPTWVISPSYKKGYTEQYGTGDPVT 406
 Db 456 t-----sfsdvpssfitgystgvt 476
 RESULT 9
 AAY58161
 ID AAY58161 standard; Protein: 588 AA.
 XX AAY58161;
 XX 07-MAR-2000 (first entry)
 DT Adeno associated virus AAV5 capsid protein VP2.
 DE Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;
 KW promoter; Rep protein; capsid protein; regulation; transcription;
 KW replication; chromosomal integration; tissue tropism; cellular receptor;
 KW gene therapy; neutralising antibody; erythroid progenitor cell;
 KW transduction; cancer; genetic disease; VP2.
 XX Adeno associated virus type 5.
 OS WO9961601-A2.
 XX 02-DEC-1999.
 XX 28-MAY-1999; 99WO-US11958.
 XX 28-MAY-1998; 98US-0087029.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chlorini JA, Kotin RM;
 PI WPT; 2000-062707/05.
 DR N-PSDB; AA249213.
 XX Adeno-associated virus 5 based vectors and particles, useful for gene
 PT therapy .
 PS Claim 27; Page 80-82; 91pp; English.
 XX This sequence represents the VP3 capsid protein of
 CC adeno associated virus type 5 (AAV5). The invention relates
 CC to vectors comprising a pair of AAV5 inverted terminal repeats
 CC (ITRs) with a promoter between the ITRs. The vector may comprise
 CC the viral genome, or subregions thereof, including sequences
 CC encoding Rep proteins and capsid proteins, and is encapsidated
 CC in an AAV5 particle. The non-structural Rep proteins Rep40
 CC (AAY58163), Rep52 (AAY58168), Rep68 (AAY58164) and Rep78 (AAY58159) are
 CC involved in regulation of replication and transcription, in addition to
 CC the production of progeny genomes. Rep68 and Rep78 are also associated
 CC with the stable integration of the viral genome into human chromosomes.
 CC The three types of capsid protein VP1 (AAY58160), VP2 (AAY58161) and VP3
 CC (AAY58162) assemble to form an icosahedral capsid, and differ from each
 CC other by the use of alternative splicing and an unusual translation
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are
 CC likely to utilise distinct cellular receptors and are serologically
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for
 CC transduction of a patient who already possess neutralising antibodies
 CC either as a result of natural immunological defence or from prior
 CC exposure to AAV2 vectors. The vectors may be useful for transducing
 CC erythroid progenitor cells or cells lacking heparin sulphate
 CC proteoglycans, which is very inefficient with AAV2-based vectors.
 CC The vectors may also be useful for transducing cells with a nucleic
 CC acid of interest in order to produce cell lines that could be used to
 CC screen for agents that interact with the gene product of the nucleic
 CC acid of interest. In addition to transduction of other cell types,
 CC transduction of erythroid cells would be useful for the treatment of
 CC cancer and genetic diseases which can be corrected by bone marrow
 CC transplants using matched donors.
 XX

us-09-426-072-2.rag

Wed Dec 5 09:18:15 2001

QY 169 GYITEEPTLVLDVDEFTTFNSWFCVPGFTNPNRLCALAGTAAGHGKNDDEFLNKGIS 228
 Db 455 grya-----ntyknwf---pdpmgtrtq-----gwnlgsqgnr-----a 484
 QY 229 SKSIFEAAEKGVSWLNVDCTNGEPEPDSLFYTYVNOTSRNVVPEVNF-----QDAYLG 284
 Db 485 svsafatnr-----melegasyqvpqngmt--nnlgsntyalentmlfnsqpanp 537
 QY 285 VLPKESYINPSCCGTNTNSMHTGNVSY-----GV 315
 Db 538 tta--cylegmilitesetqpnrvaynvggmatngsstapattynlqeiavgsv 595
 QY 316 FKQIYDAIROGPOMDKTLFTYDETGFF-----DHVPPPLAVR-----PDNL 360
 Db 596 wmer--dylgpgiaw-----ipetgahfhpamgfgikhhppmmlikntpvpgni 647
 QY 361 TYTETAKNGQYTLHEDRLGGRMPTWISPYSKGYIEQVGTDPVT 406
 Db 648 t-----sfdvpsvssfitqstgqvt 668
 RESULT 11
 AAR29028
 ID AAR29028 standard; Protein; 1257 AA.
 AC AAR29028;
 DT 20-APR-1993 (first entry)
 DE Bacillus thuringiensis toxin protein 33F2.
 KW Toxin protein; ant.
 OS Bacillus thuringiensis strain PS33F2.
 PN W09220802-A.
 PD 26-NOV-1992.
 PF 22-MAY-1992; 92WO-US04316.
 PR 22-MAY-1991; 91US-0703977.
 PR 25-NOV-1991; 91US-0797645.
 PR 12-MAY-1992; 92EP-0304228.
 XX (MYCO) MYCOGEN CORP.
 PI Kennedy MK, Meier H, Payne JM, Randall JB, Uick HJ;
 DR WPI; 1992-415780/50.
 DR N-PSDB; AAQ31411.
 XX Toxin proteins isolated from Bacillus thuringiensis - for controlling
 PT ants. e.g. fire, carpenter, argentine and pharaoh ants
 XX Disclosure; Page 45; 71pp; English.
 CC Bacillus thuringiensis toxin protein 33F2 is useful as a method of
 CC biological control of ants, e.g. fire ants, carpenter ants,
 CC argentine ants and pharaoh ants, as an alternative to chemical
 CC insecticides.
 XX Sequence 1257 AA;
 QY 62 ----RSFDNILGVRROGLDNPINNPGFCNYKNASDPSSGKYCTOAKDYDSVEND----- 112
 Db 134 ainqkfdsiqktinlytv--aidndyvtak-----tqlenlnsiltsdisif 180
 QY 113 -PDHSVTGNLEFYGYTTPNNGAIASGVV-ADQSGFLNAOLNDYPKLAPPEATROVMGY 170
 Db 181 ipegyetyg-glpvyamvanahillirdaivnaeklgfskevdtthky-----ikmtihn 234
 QY 171 YTEEVPVLVDLDEFTTF--NSWFCVPGFTNPNRLCALAGTAAGHGKNDDEFLNKGIS 228
 Db 235 hteavikaflngldkfsidvnsy-----nkkany--- 264
 QY 229 SKSIFEAAEKGVSWLNVDCTNGEPEPDSLFYTYVNTSRNVVPEVNF-----QTSRSNVVPEVNF 277
 Db 265 ikgmtemvldivalvtpfodhyqkveiefttrtisspiyqvpkqnmqntsssiivpsdlf 324
 QY 278 FODAYLGVLPK--FS-----YINPSCCGT-----NTNSMHTGN 309
 Db 325 h---yggdlvklfestrtdndglakiftgirtfykspnthetyhvfysynts---sgn 378
 QY 310 VSYG-----EVEFVKQIYD-----AIROGPOMDKTLF 336
 Db 379 isrgssnpiidlnpiistcirsnykaagssvlnfkdgqgyafaqap----- 430
 QY 337 ITYDETGEGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYTLHFDRLGGRMPTWVI 388
 Db 431 -----tggawdhstfiesdgapeghkinylytspgdrtldfinvytliustptineisteki 485
 QY 389 SPY-SKKGYIEQVGTDPVTGKP 409
 Db 486 kgfpaekgyikngqimkygk 507
 RESULT 12
 AAR28811
 ID AAR28811 standard; Protein; 1257 AA.
 AC AAR28811;
 DT 30-MAR-1993 (first entry)
 DE BT toxin 33F2.
 KW nematode worms; nematocides; nematocidal toxin; agriculture; plants;
 KW crops; pests; Cryv proteins.
 OS Bacillus Thuringiensis.
 PN W09219739-A.
 PD 12-NOV-1992.
 PF 01-MAY-1992; 92WO-US03624.
 PR 03-MAY-1991; 91US-0693018.
 PR 31-JAN-1992; 92US-0830050.
 PR 23-APR-1992; 92US-0871510.
 XX (MYCO) MYCOGEN CORP.
 PI Foncerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;
 DR WPI; 1992-398866/48.
 DR N-PSDB; AAQ30937.
 XX New genes and toxins against nematodes - obt'd. from Bacillus
 PT Thuringiensis isolates with nematocidal activity
 XX Claim 1(b) ; Page 54; 77pp; English.
 CC This sequence represents the Bacillus thuringiensis delta-endotoxin
 CC 33F2 having nematocidal activity. It was decoded from the appropriate
 CC DNA. Intact cells expressing this toxin may be used in nematocidal

QY 169 GYITEEPTLVLDVDEFTTFNSWFCVPGFTNPNRLCALAGTAAGHGKNDDEFLNKGIS 228
 Db 455 grya-----ntyknwf---pdpmgtrtq-----gwnlgsqgnr-----a 484
 QY 229 SKSIFEAAEKGVSWLNVDCTNGEPEPDSLFYTYVNOTSRNVVPEVNF-----QDAYLG 284
 Db 485 svsafatnr-----melegasyqvpqngmt--nnlgsntyalentmlfnsqpanp 537
 QY 285 VLPKESYINPSCCGTNTNSMHTGNVSY-----GV 315
 Db 538 tta--cylegmilitesetqpnrvaynvggmatngsstapattynlqeiavgsv 595
 QY 316 FKQIYDAIROGPOMDKTLFTYDETGFF-----DHVPPPLAVR-----PDNL 360
 Db 596 wmer--dylgpgiaw-----ipetgahfhpamgfgikhhppmmlikntpvpgni 647
 QY 361 TYTETAKNGQYTLHEDRLGGRMPTWISPYSKGYIEQVGTDPVT 406
 Db 648 t-----sfdvpsvssfitqstgqvt 668
 RESULT 11
 AAR29028
 ID AAR29028 standard; Protein; 1257 AA.
 AC AAR29028;
 DT 20-APR-1993 (first entry)
 DE Bacillus thuringiensis toxin protein 33F2.
 KW Toxin protein; ant.
 OS Bacillus thuringiensis strain PS33F2.
 PN W09220802-A.
 PD 26-NOV-1992.
 PF 22-MAY-1992; 92WO-US04316.
 PR 22-MAY-1991; 91US-0703977.
 PR 25-NOV-1991; 91US-0797645.
 PR 12-MAY-1992; 92EP-0304228.
 XX (MYCO) MYCOGEN CORP.
 PI Kennedy MK, Meier H, Payne JM, Randall JB, Uick HJ;
 DR WPI; 1992-415780/50.
 DR N-PSDB; AAQ31411.
 XX Toxin proteins isolated from Bacillus thuringiensis - for controlling
 PT ants. e.g. fire, carpenter, argentine and pharaoh ants
 XX Disclosure; Page 45; 71pp; English.
 CC Bacillus thuringiensis toxin protein 33F2 is useful as a method of
 CC biological control of ants, e.g. fire ants, carpenter ants,
 CC argentine ants and pharaoh ants, as an alternative to chemical
 CC insecticides.
 XX Sequence 1257 AA;
 QY 62 ----RSFDNILGVRROGLDNPINNPGFCNYKNASDPSSGKYCTOAKDYDSVEND----- 112
 Db 134 ainqkfdsiqktinlytv--aidndyvtak-----tqlenlnsiltsdisif 180
 QY 113 -PDHSVTGNLEFYGYTTPNNGAIASGVV-ADQSGFLNAOLNDYPKLAPPEATROVMGY 170
 Db 181 ipegyetyg-glpvyamvanahillirdaivnaeklgfskevdtthky-----ikmtihn 234
 QY 171 YTEEVPVLVDLDEFTTF--NSWFCVPGFTNPNRLCALAGTAAGHGKNDDEFLNKGIS 228
 Db 235 hteavikaflngldkfsidvnsy-----nkkany--- 264
 QY 229 SKSIFEAAEKGVSWLNVDCTNGEPEPDSLFYTYVNTSRNVVPEVNF-----QTSRSNVVPEVNF 277
 Db 265 ikgmtemvldivalvtpfodhyqkveiefttrtisspiyqvpkqnmqntsssiivpsdlf 324
 QY 278 FODAYLGVLPK--FS-----YINPSCCGT-----NTNSMHTGN 309
 Db 325 h---yggdlvklfestrtdndglakiftgirtfykspnthetyhvfysynts---sgn 378
 QY 310 VSYG-----EVEFVKQIYD-----AIROGPOMDKTLF 336
 Db 379 isrgssnpiidlnpiistcirsnykaagssvlnfkdgqgyafaqap----- 430
 QY 337 ITYDETGEGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYTLHFDRLGGRMPTWVI 388
 Db 431 -----tggawdhstfiesdgapeghkinylytspgdrtldfinvytliustptineisteki 485
 QY 389 SPY-SKKGYIEQVGTDPVTGKP 409
 Db 486 kgfpaekgyikngqimkygk 507
 RESULT 12
 AAR28811
 ID AAR28811 standard; Protein; 1257 AA.
 AC AAR28811;
 DT 30-MAR-1993 (first entry)
 DE BT toxin 33F2.
 KW nematode worms; nematocides; nematocidal toxin; agriculture; plants;
 KW crops; pests; Cryv proteins.
 OS Bacillus Thuringiensis.
 PN W09219739-A.
 PD 12-NOV-1992.
 PF 01-MAY-1992; 92WO-US03624.
 PR 03-MAY-1991; 91US-0693018.
 PR 31-JAN-1992; 92US-0830050.
 PR 23-APR-1992; 92US-0871510.
 XX (MYCO) MYCOGEN CORP.
 PI Foncerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;
 DR WPI; 1992-398866/48.
 DR N-PSDB; AAQ30937.
 XX New genes and toxins against nematodes - obt'd. from Bacillus
 PT Thuringiensis isolates with nematocidal activity
 XX Claim 1(b) ; Page 54; 77pp; English.
 CC This sequence represents the Bacillus thuringiensis delta-endotoxin
 CC 33F2 having nematocidal activity. It was decoded from the appropriate
 CC DNA. Intact cells expressing this toxin may be used in nematocidal

Query Match 4.38; Score 107; DB 13; Length 1257;
 Best Local Similarity 19.38; Pred. No. 2.8; Mismatches 166; Gaps 23;
 Matches 97; Conservative

QY 6 LITGLGLLASLG-LASPVTSVTSVREAPFGYKPSKESIKENLKDKNVIVWLLLEN--- 61
 Db 74 visvltlpevgtrvaasatlvsvfwkfgdkpknaklfeekpqiealqgditnyqd 133

[illegible]

XX

XX

us-09-426-072-2.rag

Wed Dec 5 09:18:15 2001

biological control agent; transgenic plant; crop improvement.
 Bacillus thuringiensis (isolate 33F2).
 US5262158-A.
 16-NOV-1993.
 30-APR-1991; 91US-0693210.
 30-APR-1992; 92US-0876280.
 30-APR-1991; 91US-0693210.
 13-SEP-1991; 91US-0759248.
 30-SEP-1991; 91US-0768141.
 (MYCO) MYCOGEN CORP.
 Bagley AL, Cannon RJC, Payne JM;
 WPI; 1993-377387/47.
 N-PSDB; AA051689.
 Controlling acaride pests e.g. two spotted spider mite - utilising
 Bacillus thuringiensis delta-toxins cloned into microbe hosts.
 Disclosure; : 42pp; English.
 DNA encoding the insecticidal toxin can be cloned into baculo
 viruses and transferred to other host microbes, preferably E. coli
 CC NM522(pMYC 2316) NRRL B-18785, to control acaride pests, or to
 CC plants which become resistant to the acaricide pests.
 CC Specifically, the two-spotted spider mite (Tetranychus urticae) is
 CC controlled.
 XX
 SQ Sequence 1257 AA;
 Query Match 4.3%; Score 107; DB 14; Length 1257;
 Best Local Similarity 19.3%; Pred. No. 2.8;
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;
 QY 6 LLTGLGLLASLG-LASPVTSYTSVREAPFGYKPGSKESIKDKVENIWLILEN--- 61
 Db 74 vlsvltiiepgvtvasaastivsfwkpikfgdkpnaknifeekpkqiealigqditnyqd 133
 QY 62 ----RSPDNLGVRROGLDNPINNGPFCNKNASDPSSGKYCTQAKDYDSVFN--- 112
 Db 134 ainqkfdsktlnlytv--aidndyvtak-----tqleninsltsdisif 180
 QY 113 -PDHSGVTGNLFEYGYTTPNNGATASGVV-ADQSGFLNAQLNDYPKLAPEATQVWGY 170
 Db 181 lpegyetg-glpymavmanahillidavnaekigfsdkvethky-----ikmtih 234
 QY 171 YTEEEVPTLVLDVDEFTTF--NSWFSCVPGPTNPRLCALAGTAAGHGKNDDDFLNIGIS 228
 Db 235 hteavikafngldkfaldvnsy-----nknany--- 264
 QY 229 SKSIFEAAKEGVMNLNYDGTNGEPEPDSLFYTYN-----QTSRNVVPVENF 277
 Db 265 lkgtentvldvalwptfdpdyhkyveleftrtisspiyqvpkmmqntsssiyvspdlf 324
 QY 278 QDAYLVLPK--FS-----YINPSCCGT-----NTNSMHPTGN 309
 Db 325 h---yqgdvklfstrtdndgklakiftgirtfyspnthetyhvdfsyntqs---sgn 378
 QY 310 VSYG-----EVFKQIYD-----AIROGPQWOKTLIF 336
 Db 379 isrgssnpipidlnnplictrnsfyalagssvlvfkdgtdggyafadap----- 430
 QY 337 IYDETGGFVDH--VPPPLAVRPDNLJYTTETAKNG-----QKYLHFDRLGRMPTWVI 388
 Db 431 ----tggawdhfiesdgapeghklnlytspgdclrdfnvtyllstptinelsteki 485

QY 389 SPY-SKKGYIEQYCTDPTVTKP 409
 Db 486 kgfpaekyiknggimkyygkp 507
 RESULT 15
 AAW13886
 ID AAW13886 standard; protein; 1257 AA.
 XX
 AC AAW13886;
 XX
 DT 14-MAY-1997 (first entry)
 XX
 DE 33F2 toxin.
 XX
 KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;
 KW biological control; Monomorium pharaonis; delta-endotoxin; lepidoptera;
 KW Insect.
 XX
 OS Bacillus thuringiensis isolate PS33F2.
 XX
 PN US5596071-A.
 XX
 PD 21-JAN-1997.
 XX
 PF 22-MAY-1991; 91US-0703977.
 XX
 PR 24-NOV-1993; 93US-0158232.
 PR 22-MAY-1991; 91US-0703977.
 PR 25-NOV-1991; 91US-0797645.
 PR 22-MAY-1992; 92US-0887980.
 XX
 PA (MYCO) MYCOGEN CORP.
 XX
 PI Foncerrada L, Fu J, Kennedy MK, Meier H, Payne JM;
 PI Randall JB, Schnepf HE, Schwab GE, Ulick HJ;
 DR
 DR WPI; 1997-107615/10.
 DR N-PSDB; AAT60071.
 XX
 PT Bacillus thuringiensis toxin - active against hymenopteran pests
 XX
 PS Disclosure; Column 57-64; 64pp; English.
 XX
 CC This sequence represents the 33F2 toxin isolated from the Bacillus
 CC thuringiensis (B.t.) isolate PS33F2. B.t. is a gram-positive, spore
 CC forming, soil bacterium, characterised by parasporal crystalline protein
 CC inclusions. These proteins can be highly toxic to pests, and have been
 CC used to produce insect resistant plants. The previously isolated B.t.
 CC delta-endotoxins were mainly active against lepidopteran insects,
 CC however the proteins of the invention are active against hymenopteran
 CC insects. This sequence is an example of a toxin of the invention, for
 CC which the sequences shown in AAW13888 and AAW13871 represent the generic
 CC formulae. As the toxins of the invention are active against hymenopteran
 CC pests, they can be used for the biological control of ants, particularly
 CC pharaoh ants (Monomorium pharaonis).
 XX
 SQ Sequence 1257 AA;
 Query Match 4.3%; Score 107; DB 18; Length 1257;
 Best Local Similarity 19.3%; Pred. No. 2.8; Mismatches 166; Indels 166; Gaps 23;
 Matches 97; Conservative 73;
 QY 6 LLTGLGLLASLG-LASPVTSYTSVREAPFGYKPGSKESIKDKVENIWLILEN--- 61
 Db 74 vlsvltiiepgvtvasaastivsfwkpikfgdkpnaknifeekpkqiealigqditnyqd 133
 QY 62 ----RSPDNLGVRROGLDNPINNGPFCNKNASDPSSGKYCTQAKDYDSVFN--- 112
 Db 134 ainqkfdsktlnlytv--aidndyvtak-----tqleninsltsdisif 180
 QY 113 -PDHSGVTGNLFEYGYTTPNNGATASGVV-ADQSGFLNAQLNDYPKLAPEATQVWGY 170

181	ipegyetg-gipyyamvanahlllrdalvnaeklgfskdevdthky-----ikmthn	234
171	ITVEEYPTLVLDVDEFTTF--NSWFCVPGPTPNRLCALAGTAGCKNDLFLNYGIS	228
235	hteavikafngldkfksldvnsy-----nkkany---	264
229	SKSIFPAANEKGVSWLNYDCTGCEFPDPSLFYVYV-----QTSRSVNVDPNF	277
265	lkgmtemvldivalwlpfdpbdhyqveleftrtisslyqpvpkmmqntssslvpsdlf	324
278	FQDAYLGVLPK--FS-----YINPSCCGT-----NTNSMPTGN	309
325	h---yggdlvklefstrdndglakifgirnfykspthetyhvfdsyntgs---sgn	378
310	VSVG-----EYFVKQIYD-----AIRGGPQWDKTLFL	336
379	isrgsnnpipdlnopliistcirsfnfkalgagssvlnfkdgtygaafagp-----	430
337	ITVDETFGFDYH--VPPELAVRPDLNLTVTETAKNG-----QKYVTHFDLGRGMP	388
431	-----tggawdhsfiesdgapeghklnlytcspgdtrdfinvfllstptinelsteki	485
389	SPY-SKKGYIEQYGTDPVTCGP	409
486	kgfpaeqgyiknggimkygk	507

Search completed: December 4, 2001, 15:07:35
Job time: 36 sec

Db 289 REVKEVETLRSSPQWKEMALLITYDEHGGFYDHPVTPVKGVNPDGII-----GPDFF 342

QY 373 TLHEDRLGGMPWTWISPSKGYI--EQYGDTPVTGKPAPYSA---TSVLKTLGLWMDI 427

Db 343 YGFGDGLGVKVPFLISSWIEKGVHIEPEG-----PTPHSQFHESSIPATVKKLFNL 395

QY 428 ED--FTPRVAHSPFDHL--IGTTLREDAP 453

Db 396 KSHFTLRDAWAGTEKVFRIKRDSPQDCP 425

RESULT 2

T46205

hypothetical protein T8P19.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46205

R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; S

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223008

A:Accession: T46205

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-462 <CHO>

A:Cross-references: EMBL:AL133315

A:Experimental source: cultivar Columbia; BAC clone T8P19

C:Genetics:

A:Map position: 3

A:Introns: 139/2; 307/1

A:Note: T8P19.120

Query Match 17.4%; Score 436; DB 2; Length 462;

Best Local Similarity 32.0%; Pred. No. 3.1e-24;

Matches 125; Conservative 61; Mismatches 165; Indels 40; Gaps 13;

QY 78 NPINN---GPFCKYKNASDPSSGKYCTQAKVDVNDPDPDHSVTGNLLEFYGYTPNNGA 134

Db 6 NFTNGVTGECN----PVPNSTQITCTFSDAEFVDPDPGHSFEAVEQQVFGS---GPGQ 58

QY 135 TASGVVADSGFLNAQLNDPKLAPAEATQVMGYITEEVPFLVDLDEFTTNSWFS 194

Db 59 IPS-----MNGVEQALS-----MPCNLSETVMKGFPEAPVPIAEVKEAFVDRWFS 107

QY 195 CVPGPTPNRLCALAGTAGHGKNDDEFNLNGYISKSIFFAANKGVSWLNDGTNGE 254

Db 108 SIPCPTQPNRLFVSATSHGSTSHVKKLAQGYPKQTFIDSLHSNDIDF---GIYFQNI 163

QY 255 PDSLEFFVYVQTSR-SNVVPVE-NFQDAYLVLPKFSYINP---SCCGTNTNSMHTGN 309

Db 164 PFTFLYRLQLKXIFNLHQYDLKFKDAKGLPSLVIEPRYFDLGLPANDDHPSHD 223

QY 310 VSYGEVFKQIYDAIROGPQWDKTLLEITYDETGFGYDHPVPPPLAVRPNLTYTTAKNG 369

Db 224 VANGOKLVKEVEALRSSPQNNETLLIITYDEHGGFYDHVKTPVVGIPNPDGNTGPAGF 283

QY 370 QKYLTLHEDRLGGMPWTWISPSKGYIEQYGDTPVTGKPAPYSAVSLKTLGLWMDIED 429

Db 284 FK----FDRGLGVKVPFMVSEWIKQGVSEAKGPT--ESSEVEHSSIPATIKKLFNLSS 337

QY 430 --FTPRVAHSPFDHLIG--TTLREDAPIAL 456

Db 338 NFLTHRDAAWATEFVWSHLTTPRTDCPMTL 368

RESULT 3

T02648

probable phospholipase C [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein F12C20.9

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001

C:Accession: T02648; H84665

R:Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; S

submitted to the EMBL Data Library, August 1998

A:Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.

A:Reference number: Z14685

A:Accession: T02648

A>Status: translated

A:Molecule type: DNA

A:Residues: 1-514 <ROU>

A:Cross-references: EMBL:AC005168; NID:g3426033; PID:g3426039

A:Experimental source: cultivar Columbia

R:Lib, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487

A:Accession: H84665

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 <STO>

A:Cross-references: GB:AE002093; NID:g3426039; PIDN:AC32238.1; GSPDB:GN00139

C:Genetics:

A:Gene: F12C20.9; At2g26870

A:Map position: 2

A:Introns: 156/2; 280/3

Query Match 17.4%; Score 435.5; DB 2; Length 514;

Best Local Similarity 29.8%; Pred. No. 4e-24;

Matches 130; Conservative 79; Mismatches 162; Indels 65; Gaps 17;

QY 51 VENIVMLILENSFDNIIIGVRRGLDNPIN--NGPFCNYKNASDPSSGKYCTQAKYDS 108

Db 27 IKTIIVVMENRSFDMGLNMKK--LNPEINGVDGSESNPVSVDPSRRKIFGSGSH-Y 83

QY 109 VFNDPDSHYTGNLEFYCTY-----TPNNGAIASGVVADSGFLNAQLNDYKLAPEE 162

Db 84 VDDPGHSGQATREOVFGSNDTSMDDPPMNGFVQQA-YSEDPGNSMA----- 130

QY 163 ATROVMGYTEBEVPLVDLDEFTTNSWFSVCEGPTPNRLCALAGTAGHGKNDDEF 222

Db 131 ---SVMNGEPKVPYKSLVSEFAVDFRWFASVPSSTQPNRMFVHSGTSAGATSNPNIS 187

QY 223 LNYGISKSIFFAANKGVSWLNDGTNGEFPEDSLPFTVYNOT-----SRSNVVPVE 275

Db 188 LARGYQRTIFDNLDEEFSE---GIYQNIPLAVLFQSLKLYVFKFSYGN----- 238

QY 276 NFQDAYLVLPKFSYINPSCCGT---NTNSMHTGNVSYGEVFPVQIYDAIROGPQWDK 332

Db 239 SFKDHAQOKLPAYTIEQRYMTLLEPASDDHPSHDYQOGKPIKEVYETLRASPOWNE 298

QY 333 TLLFIYDETGFGYDHPVPPPL--AVRPDNLTYTTAKNGOKYTLHEDRLGGMPWTWISP 390

Db 299 TLLIITYDEHGGFYDHPVTPVRNVPSPDGIV-----GPDPLFQPNRLGIRVPTIAVSP 352

QY 391 YSKGYIEQYGDTPVTGKPAP---YSATSVLKLGLWMDIED--FTPRVAHSPFDHL-- 443

Db 353 WIEKGV-VHGNP---GSPFSESEHSSIPATVKKLFNLSSPFLTKRDEWAGTFENILQ 408

QY 444 IGTTLREDAPIALKTP 459

Db 409 IRKEPRDTCPELTPEP 424

RESULT 4

F70662

probable plcC protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: F70662

R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: F70662
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <COL>
A:Cross-references: GB:283860; GB:AL123456; NID:g3261681; PIDN:CAB06146.1; PID:g1781256
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: plcB

Query Match 13.0%; Score 324.5; DB 2; Length 508;
Best Local Similarity 26.0%; Pred. No. 5.1e-16;
Matches 140; Conservative 61; Mismatches 203; Indels 135; Gaps 24;
QY 10 LGLLASGLASPTSEYTSVREAPFGYKPGKSEINLKDKNVENIWLILNRSFDNILG 69
DB 7 LAKAAGAGAAVLTDWAAPVIEKAYGAGPCS---GHUTD-IEHIVLCLOENRSDHYFG 61
QY 70 GVR-OGDNPINGPCFNKASDPSSGKYCTQAKDYDSV-----F 110
DB 62 TLSAVDGFDTPT---PLFQKQWNP-----TQALDPTGTTLPYRINTGGPGNGGECV 112
QY 111 NDPHSVTGNLEFYGTTPNNGAIASGKVVADQSGFLNAQLANDYKPLAPEATROVMGY 170
DB 113 NDPDQWIAAHLW-----NGGA-----NDGWLPAQ---ARTSRVANTPVVMGY 153
QY 171 YTEEVPTLVDLVEFTTFNSWFSVCPGTPNRLCALAGTAAGH-KNDQDFLNYGSS 229
DB 154 YARDIPPIHYLLADFTTCDQYFSLGGTNPRLYISATVNDQGGPOIVEPALOP 213
QY 230 KSIP-----EAANEKGVSW-----LNYDGTNGEF-----EPDSLFFTYN 264
DB 214 KLTFWRIMPQNLSDAGISWKNVSKLLGLNDTSLSRNGYVSGFSKQAADPRS----- 266
QY 265 QTSRSNVVPE--NFFQDAYLGVLPKFSYINPSCGTNTNSMHTGNVSYGEVFKQIYD 322
DB 267 DLARYGAPAYPWFIRDVINNLQVSWVP---LNVSEHSPFVAVGATVNLIR 322
QY 323 AIROGPO-WDKTLFTTYDETGFGYDHPPLAVRPNMLTYETAKNGOKYTLHDDR--- 378
DB 323 VLLRNPVWEKTAIIAYDEHGFHDVHTPLTA--PEG-TPGEWIPN---SVDIDKYDG 375
QY 379 -----LGGRMPTWISPYSKGYIEQYGTDPVT-----GKAPYSATSVLKT 420
DB 376 SGGIRGPIGLGFRVCFVISPYSRGGLVHVDHEDHTSOLQLIGKRGVPPNLTWRASV 435
QY 421 LGYLWLDIEDF-TPRVAHSPFSDH-----LIGTTLREDAPALKTPTFSY 464
DB 436 TGDMTSAFNAAPPDPPNLDHVRQLPKVAKCPNPNVVLGFLNEGLPYRVPYPTTPV 494

RESULT 5
G70662
probable plcB protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70662
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: G70662
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-512 <COL>
A:Cross-references: GB:283860; GB:AL123456; NID:g3261681; PIDN:CAB06147.1; PID:g1781257
A:Experimental source: strain H37Rv

C:Genetics:
A:Gene: plcB
Query Match 11.5%; Score 288; DB 2; Length 512;
Best Local Similarity 26.1%; Pred. No. 2.4e-13;
Matches 141; Conservative 60; Mismatches 184; Indels 156; Gaps 29;
QY 14 ASGLASPTSEYTSVREAPFGYKPGKSEINLKDKNVENIWLILNRSFDNILGVR- 72
DB 18 AFMSLAGPTIEKAYGA-----GCPG-----HLTD-IEHIVLLMOENRSDHYFGTSLD 65
QY 73 RQGLD-----NDPNNGPCFNKASDPSSGKYCTQAKDYDSV-----F 110
DB 66 TRGDDTTTPPVFAQSGWNPMT-----QAQDPAG---VTLPYRFDTTRGPLVAGECV 114
QY 111 NDPHSVTGNLEFYGTTPNNGAIASGKVVADQSGFLNAQLANDYKPLAPEATROVMGY 170
DB 115 NDPDHSWIGMHNSW-----NGGA-----NDNWLPAQVFPFSLQGNVPVT---MGF 156
QY 171 YTEEVPTLVDLVEFTTFNSWFSVCPGTPNRLCALA-----CTAAGH---GKNDDD 221
DB 157 YTRRDLPIHYLLADFTTCDQYFCSLLGCTTPNRLYNSAWIDPDGTDGPPVLIENIOP 216
QY 222 FLAYGSSKSEIFEAANEKGVSW-----LNYDGTNGEF-----EPDSLFFTY 262
DB 217 LQHY--SWRIMPENLEDGWSKVKYONKLLGALNTVVGYNGLVDFKQAADPRS----- 269
QY 263 VNQTSRSNVVPE--NFFQDAYLGVLPKFSYINPSCGTNTNSMHTGNVSYGEVFKQI 320
DB 270 --NLARFGISPTYPDLDAADVRNRLPKVSWVLPF---LLSEHPAFPPVNVGAV---AI 320
QY 321 YDAIR--OGPO-WDKTLFTTYDETGFGYDHPVP---LAVRPDNLTYTETAK 367
DB 321 VDAURLILSNPAVWEKTAIIYNDENGDFHDVHTPPPTGCTGCEFTVTPD--IDSVGS 378
QY 368 NGOKYTLHFDRLGRMPTWISPYSKGYIEQYGTDPVT-----GKAPYSATSV 417
DB 379 GGIIRGPI--GLGFRVCPVISPYS-RGPLVMVHDFDHTSTLKLIRARFGVPPNLTAMR 434
QY 418 LKTYGLWLDIEDF-TPRVAHSPFSDH-----LIGTTLREDAPALKTPTPH 460
DB 435 DATVGDMTSTFNAAPPNPKPNLDHRLNALPKLQCPVNAVLGTVTKTAIYRVPFPQ 494
QY 461 T 461
DB 495 S 495

RESULT 6
A26391
phospholipase C (EC 3.1.4.3) - Pseudomonas aeruginosa
N:Alternate names: lecithinase C; lipophosphodiesterase I
C:Species: Pseudomonas aeruginosa
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 08-Oct-1999
C:Accession: A26391
R:Pritchard, A.E.; Vasil, M.L.
J. Bacteriol. 167, 291-298, 1986
A:Title: Nucleotide sequence and expression of a phosphate-regulated gene encoding a
A:Reference number: A26391; MUID:86250607
A:Accession: A26391
A:Molecule type: DNA
A:Residues: 1-730 <PRI>
A:Cross-references: GB:M13047; NID:g151492; PIDN:AAA25966.1; PID:g151493
C:Keywords: phosphoric diester hydrolase
Query Match 11.5%; Score 287.5; DB 2; Length 730;
Best Local Similarity 23.9%; Pred. No. 4.3e-13;
Matches 119; Conservative 62; Mismatches 156; Indels 161; Gaps 22;
QY 51 VENVIVLILNRSFDNILG---GVRQGLDNPINNGPFCFNKASDPSSGK---YCTQAK 104
DB 104 YCTQAK 104

Db 51 VOHVILMOENRSDHYFGLNGVR--GENDP-----RALKRODGKPVWY---Q 94
QY 105 DYDSVNDPDHSTVGNLFYGYTTPNNGAIASCKVADQS-----CFLNAQLND--- 154
Db 95 NYKFEF-PYHWDT-----KVTSAQWSSQNHWSAFHAINQGRNCKWM 138
QY 155 ---YPKLAPEEATQVMGYTTEEVPTLDVDEFTTFNSWFCVPGPTNPNRLCALAGT 211
Db 139 AVQVP-----EAMGYFKRGDIPYYALADAPTLC EAYHQSMMGPTNPNRLYHMSGR 189
QY 212 AAGHGKNDLFLNYGSSKSF-----EAAKEGVSMLNDGTNGEFPDLSFF 260
Db 190 AAPSGDGRVHIGNDMGDCITIGASGIVDWTTPYERLSAAGVDRVYQ---EGYRSSSLWY 247
QY 261 TYVNOTSRNSVVPVNF-----FQDAYL-----GVLP 287
Db 248 LYVDAYWKYRLOEQNNYDCNALAWFRNFKNAPRDSDLWQRAMLAGVQDLRKDVQENTLP 307
QY 288 KFSYINPSCCGTNTSMHPTGNVSYGEVFKQIYDAIROGPQ-WDKTLLFFITDETGGFY 346
Db 190 AAPSGDGRVHIGNDMGDCITIGASGIVDWTTPYERLSAAGVDRVYQ---EGYRSSSLWY 247
QY 261 TYVNOTSRNSVVPVNF-----FQDAYL-----GVLP 287
Db 248 LYVDAYWKYRLOEQNNYDCNALAWFRNFKNAPRDSDLWQRAMLAGVQDLRKDVQENTLP 307
QY 288 KFSYINPSCCGTNTSMHPTGNVSYGEVFKQIYDAIROGPQ-WDKTLLFFITDETGGFY 346
Db 308 QVSWIVAPYCYE---HPWGFSGEYVYTRVLEALTSNPEVWARTFILNYDEGDFY 363
QY 347 DHVPPPLAVRPNLYTETAKNGQKYLHFR-----LGGRMPTWVISPYSKKGYIEQYG 401
Db 364 DHASAPVPPWKDGVGLSTVSTAGE---IEVSSGLPIGLGHRVPLIAISPSWKGKGV--- 416
QY 402 TDPVTGKPAPIYATSVLTKL-----GYLWDIEDF-----TPRV 434
Db 417 -----SAEVFDHTSVLRFLERFGLVEENISPWRAVCGDLTSLFDFOGAGDTQVAPDL 470
QY 435 AHSFSDHLLIGTLTRED 452
Db 471 TNVPQSD-----ARKEDA 483

RESULT 7

B83540
hemolytic phospholipase C precursor PA0844 [imported] - Pseudomonas aeruginosa (strain H
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83540
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lin,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: B83540
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-730 <STO>
A:Cross-references: GB:AE004519; GB:AE0046736; PIDN:AAG04233.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: plCH; PA0844

Query Match 11.58; Score 287.5; DB 2; Length 730;
Best Local Similarity 23.98; Pred. No. 4.3e-13;
Matches 119; Conservative 62; Mismatches 156; Indels 161; Gaps 22;

QY 51 VENTVWILLENRSDHYFGLNGVR--GENDP-----RALKRODGKPVWY---Q 94
Db 51 VOHVILMOENRSDHYFGLNGVR--GENDP-----RALKRODGKPVWY---Q 94
QY 105 DYDSVNDPDHSTVGNLFYGYTTPNNGAIASCKVADQS-----CFLNAQLND--- 154
Db 95 NYKFEF-PYHWDT-----KVTSAQWSSQNHWSAFHAINQGRNCKWM 138
QY 155 ---YPKLAPEEATQVMGYTTEEVPTLDVDEFTTFNSWFCVPGPTNPNRLCALAGT 211
Db 139 AVQVP-----EAMGYFKRGDIPYYALADAPTLC EAYHQSMMGPTNPNRLYHMSGR 189

QY 212 AAGHGKNDLFLNYGSSKSF-----EAAKEGVSMLNDGTNGEFPDLSFF 260
Db 190 AAPSGDGRVHIGNDMGDCITIGASGIVDWTTPYERLSAAGVDRVYQ---EGYRSSSLWY 247
QY 261 TYVNOTSRNSVVPVNF-----FQDAYL-----GVLP 287
Db 248 LYVDAYWKYRLOEQNNYDCNALAWFRNFKNAPRDSDLWQRAMLAGVQDLRKDVQENTLP 307
QY 288 KFSYINPSCCGTNTSMHPTGNVSYGEVFKQIYDAIROGPQ-WDKTLLFFITDETGGFY 346
Db 308 QVSWIVAPYCYE---HPWGFSGEYVYTRVLEALTSNPEVWARTFILNYDEGDFY 363
QY 347 DHVPPPLAVRPNLYTETAKNGQKYLHFR-----LGGRMPTWVISPYSKKGYIEQYG 401
Db 364 DHASAPVPPWKDGVGLSTVSTAGE---IEVSSGLPIGLGHRVPLIAISPSWKGKGV--- 416
QY 402 TDPVTGKPAPIYATSVLTKL-----GYLWDIEDF-----TPRV 434
Db 417 -----SAEVFDHTSVLRFLERFGLVEENISPWRAVCGDLTSLFDFOGAGDTQVAPDL 470
QY 435 AHSFSDHLLIGTLTRED 452
Db 471 TNVPQSD-----ARKEDA 483

RESULT 8

H70662
probable pica protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70662
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: H70662
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-512 <COL>
A:Cross-references: GB:283860; GB:AL123456; NID:g3261681; PIDN:CAB06148.1; PID:g17812
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pica

Query Match 11.44; Score 286; DB 2; Length 512;
Best Local Similarity 28.68; Pred. No. 3.3e-13;
Matches 128; Conservative 40; Mismatches 161; Indels 118; Gaps 20;

QY 7 LTGLGLLASL-GLASPVTSYTSVREAPPGYKPGSKESIEINKDKVENIWLILNRSFD 65
Db 10 LTGAGAAFLMDWAAPVIEKAYGA-----GPCPG-----HLTD-IEHIVLLMOENRSED 57
QY 66 NILGQVRRGLDNLNNGPFCNKNASDPSSGKY-----CTQAKDYDSV----- 109
Db 58 HYFETLSS-----TNGFNASPAFCQMGWNPWTOALDPAGVTIFRDLTTRGP 105
QY 110 -----FNDDPHSVTGNLNFYGYTTPNNGAIASCKVADQSGLFLNAQLNDYPKLAPEEA 163
Db 106 FLDEGVNDPEHQQWVGHAW-----NGGA-----ND--NWLPAQA 139
QY 164 TRQ-----VMGYTTEEVPTLDVDEFTTFNSWFCVPGPTNPNRLCAL-----AG 210
Db 140 TTRAGYVPLVINGYTTQDPIHLLADFTTICDGYHCSLLTGLTPNRLYLSANIDPAG 199
QY 211 TAAGHKNDLFLNYGSSKSF-----EAAKEGVSMLNDGTNGEFPDLSFFITVNTSR 268
Db 200 TDGPGQVPEPGFLPLQOQFSWRIMPENLEDAGVSKVQNGKLGRTINTPTISNGLVQAFR 259
QY 269 SNVVPVNF-----FFQDAYLGLVLPKFSYINPSCCGTNTSMHPTGNVSYGEV 315

us-09-426-072-2.rpr

Wed Dec 5 09:18:18 2001

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: E83230
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-692 <STO>
A:Cross-references: GB:AE004754; GB:AE00491; NID:g9949446; PIDN:AAG06707.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: plcN; PA3319

Query Match 10.6%; Score 264; DB 2; Length 692;
Best Local Similarity 25.8%; Pred. No. 2.1e-11;
Matches 108; Conservative 59; Mismatches 163; Indels 88; Gaps 19;

QY 46 NLKDKVENIVWLILENRSFDNIIIG---GVRROG---LDNPNNGPFCNKNASDPSSGKY-99
DB 44 NLKD-VEHVILMQENRSFDHYFTGLKGVGRGDRMAIPLDQGRVWHQKSGKEILPYH 102
QY 100 ----CTQAKDYSVNDPDHSVTGNLLEYGTYTPNNGATASCKVVDQSGFLNAQLNDY 155
DB 103 FDTSTTSQORVDG-----TPHTWPD-----QAANNEGRMD--133
QY 156 PKLAPEEATROVMGYTEEEVPTLVLDVEFTTNSWFSVCVPGTPNPNRICALAGT---A 212
DB 134 -KWLPAK-TERSLGYKQDIAFOFAMANAFTICDAYHCSFGQGTNPNRFLMTGTNDPL 191
QY 213 AGHG---KNDDDF---LNGVSGSKSIFEAANEKGVSWLNYDGTNGEREPSDL---FFTYV 263
DB 192 GQHGPGVPTTNDHSDNGVEGQYTWTPYPERLOAGITWRVYQDMADNFSNPLIGFRQYR 251
QY 264 NOTSRNVY-----PVENFFQDAYLGVLPKFSYINPSCCGTNTNSMHP-TGNVSYGEV 315
DB 252 AAAPDPLVINGLSTWKLKALKRDVLANSLPOVSWI-----VAPAKYSEHPGSPSPINGAE 307
QY 316 FVKQIYDAIROGPO-WDKTLLFTYDETFGFDYDHPVPPPLA--VRPNLTYTET-----365
DB 308 YTSWLDALTANPEVSWKALLVHFDENGDFDHPVAPPAAPSLNKGDTLGRKTTADATLE 367
QY 366 --ARKQKTYLHFDRLGGRMPTWVISPYSKGYIEQYDTPVTKGKAPYSATSVLKT 421
DB 368 WHTKGDIRYRNOPYGLGPRVPMYVISPWSKGGWNS-----QVFDHTSVIRFL 415

RESULT 11

T35009
probable phospholipase C - *Streptomyces coelicolor* (fragment)
C:Species: *Streptomyces coelicolor*
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35009
R:Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21565
A:Accession: T35009
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-501 <SEE>
A:Cross-references: EMBL:AL079355; PIDN:CA845565.1; GSPDB:GN00070; SCOEDB:SC4C6.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC4C6.01

Query Match 8.1%; Score 202.5; DB 2; Length 501;
Best Local Similarity 26.7%; Pred. No. 4.1e-07;
Matches 63; Conservative 28; Mismatches 78; Indels 67; Gaps 8;

QY 210 GTAAGHGKNDLDFLNYGISKSIPEAANEKGVSWLNY-----DGTNGE 252

DB 260 QARADPSNLARYGIATPYCDFAADVRAANRPKVSMLVFN-----LQSEHPALPVALGAV 315
QY 316 FVKQIYDAIROGPO-WDKTLLFTYDETFGFDYDHPVPPPLA-----VRPNLTYTET 366
DB 316 SMVTALRILLSNPAWEKITALIVSYDENGDFDHPVAPPAAPSLNKGDTLGRKTTADATLE 374
QY 367 KNGQKTYLHFDRLGGRMPTWVISPYSK 393
DB 375 SGGIRGPL---GLGFRVPCIVISPYSR 398

RESULT 9

A36143
phospholipase C (EC 3.1.4.3), nonhemolytic - *Pseudomonas aeruginosa*
C:Species: *Pseudomonas aeruginosa*
C:Date: 30-Nov-1990 #sequence_revision 30-Nov-1990 #text_change 08-Oct-1999
C:Accession: A36143
R:Ostroff, R.M.; Vasil, A.L.; Vasil, M.L.
J. Bacteriol. 172, 5915-5923, 1990
A:Title: Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from *Pse*
A:Reference number: A36143; MUID:91008968
A:Accession: A36143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-692 <STO>
A:Cross-references: EMBL:M59304; NID:g151495; PIDN:AAA25968.1; PID:g151496
C:Keywords: phosphoric diester hydrolase

Query Match 10.7%; Score 267; DB 2; Length 692;
Best Local Similarity 25.2%; Pred. No. 1.3e-11;
Matches 114; Conservative 64; Mismatches 181; Indels 94; Gaps 20;

QY 11 GLASLGLASPTSYTSVREAPFGYKPGSKESTENLKDKNVENIVWLILENRSFDNIIIG-69
DB 15 GTVANVATSMPLPSSIQALAIAPHRHG-----NLKD-VEHVILMQENRSFDHYFTG 67
QY 70 --GVRROG---LDNPNNGPFCNKNASDPSSGKY-----CTQAKDYSVNDPDHSVTGN 120
DB 68 LKGVGRGDRMAIPLDQGRVWHQKSGKEILPYHFTSTTSQORVDG-----115
QY 121 NLEFYGTYTPNNGATASCKVVDQSGFLNAQLNDYPKLAPEEATROVMGYTEEEVPTLV 180
DB 116 -----TPHTWPD-----QAANNEGRMD--KWLPAK-TERSLGYKQDIAFOF 156
QY 181 DLVDEFTTNSWFSVCVPGTPNPNRICALAGT---AAGHG---KNDDDF---LNGVSGSK 230
DB 157 AMANAFTICDAYHCSFGQGTNPNRFLMTGTNDPLGQHGGEVPTTNDHSDNGVEGQYTW 216
QY 231 SIFEANEKGVSWLNYDGTNGEREPSDL---FFTYVNOTSRNVY-----PVENFFQDA 281
DB 217 TYPERLOAGITWRVYQDMADNFSNPLIGFRQYRAAPDPLVINGLSTWKLKALRDV 276
QY 282 YLGVLPKFSYINPSCCGTNTNSMHP-TGNVSYGEVFKQIYDAIROGPO-WDKTLLFTY 339
DB 277 LANSLPOVSWI-----VAPAKYSEHPGSPSPINGAEYTSWLDALTANPEVSWKALLVNF 332
QY 340 DETGGFDYDHPVPPPLA--VRPNLTYTET-----ARKQKTYLHFDRLGGRMPTWVI 388
DB 333 DENGGFDHVAAPPAAPSLNKGDTLGRKTTADATLEWHTKGDIRYRNOPYGLGARVPMYVI 392
QY 389 SPYSKGYIEQYDTPVTKGKAPYSATSVLKT 421
DB 393 SPWSKGGWNS-----QVFDHTSVIRFL 415

RESULT 10

E83230
non-hemolytic phospholipase C precursor PA3319 [imported] - *Pseudomonas aeruginosa* (stra
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83230

C:Species: Kluyveromyces marxianus var. lactis, (Candida spuerficia)
C.Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 20-Jun-2000
C.Accession: B31776
R.Chang, T.D.; Dickson, R.C.
J. Biol. Chem. 263, 16696-16703, 1988
A.Title: Primary structure of the lactose permease gene from the yeast Kluyveromyces
A.Reference number: A92683; MUID:89034156
A.Accession: B31776
A.Molecule type: DNA
A.Residues: 1-422 <GB>
A.Cross-references: NCBI:06997; NID:g2856; PIDN:CAA30054.1; PID:q1325996

Query Match	6.3%; Score 158; DB 2; Length 280;
Best Local Similarity	28.6%; Pred. No. 0.0032;
Matches	53; Conservative 21; Mismatches 67; Indels 8;
QY	238 EKGYSWLNVDGTTNGEPEPDSLFFTYVNOTS-----RSNVVP--VENFFQDAY 282
Db	1 DAGYSKWVY--RNKLTGPISSVLYIGSLVTSFKQSADPRSDLVREGVAPSPASFAADV L 58
QY	283 LGVLPKFSYINPSCCCTNTNSMHPGTGNVSGEYFVKQIYDAIROGPO-WDKTLFLTYDDE 341
Db	59 ANRLPRYSWIPNVN----LESEHPAVPAAGAFAIVNILRILLANPAVWEKTALIVSYDE 114
QY	342 TGGFYDHVVPPPLAVRPDNLITETAKNGOKYTL-HFDR-----LGRGMTPTWI 388
Db	115 NGGFEDRVDPATA-----PAGTPEGVTVPDDIDQVPGSGGIRGPIGLGFRVPCFVI 165
QY	389 SPYSK 393
Db	166 SPYSR 170
RESULT	14
JC7179	acid phosphatase (EC 3.1.3.2) Pho610 - yeast (Kluyveromyces marxianus)
N:Alternate names:	Pho610 protein
C:Species:	Kluyveromyces marxianus
C:Date:	04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
C:Accession:	JC7179
R:Yoda, K.; Ko, J. H.; Nagamatsu, T.; Lin, Y.; Kaibara, C.; Kawada, T.; Tomishige, N.;	Biosci. Biotechnol. Biochem.
A:Title:	Molecular characterization of a novel yeast cell-wall acid phosphatase clone
A:Reference number:	JC7179; PMID:20169637
A:Accession:	JC7179
A:Molecule type:	DNA
A:Residues:	1-483 <YOD>
A:Cross-references:	GB:E02615
A:Experimental source:	Strain Y-610
C:Comment:	This enzyme, highly glycosylated, is a covalently-linked cell-wall protein organic compounds.
C:Genetics:	
A:Gene:	pho610
C:Keywords:	cell wall; endoplasmic reticulum; glycoprotein; phosphoric monoester hydr
Query Match	5.2%; Score 130.5; DB 2; Length 483;
Best Local Similarity	23.4%; Pred. No. 0.071;
Matches	68; Conservative 47; Mismatches 111; Indels 65; Gaps 14;
QY	134 AIAAGKWADO-SGFLNALN-----DYPKLAPEATROVMGYITEEVPTLVDL 182
Db	48 AAASSAVTHRTISDVKAARVRVFVILENTDYDKAAEDE-----SLAWL 92
QY	183 VDEFITFNSSVCVPGTPNNRLCALAGTAGHGNDDDFLNYGISSKSTFEAANEKGSV 242

Db 93 AKOISLTNYS-LTHSEPNYLASVAGDY--FALDDDFISMPSNVSNIVDLDTKNIS 149
Qy 243 W-----LNYDGTNG-----EFEPDSLFTYVYNOTSR-SNVVPPVENFFOD 280
Db 150 WAQYQHPIPYTGFQGFYNSQETFEANDYVRKHNPILLDFDNVISDKSRLANIKSFEDFNFD 209
Qy 281 AYLGLVLPKFSYINPSCGTNTNSMPTGNVSIGEVFKQIYDAIRQGPQMDK-TLLFIT 339
Db 210 LKNKTLPOVAETPNM---TNDGHS-NIKVAGDWSKFLQPLLSDDYFMKDTLVLLTF 264
Qy 340 D--ETGGFDVHPPPL--AVRPDNLTYTETAKNGKYTLHFDRLGSRPTW 386
Db 265 DENETGKIKNVFSILGGVDPDLKGT-----SDOTFYDHSOLATVEANW 311

RESULT 15
T35551
probable phospholipase C - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35551
R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21582
A:Accession: T35551
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-221 <SEE>
A:Cross-references: EMBL:AL096825; PIDN:CAH4887.1; GSPDB:GN00070; SCODEB:SC6G3.07
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SC6G3.07

Query Match 5.1% Score 128; DB 2; Length 221;
Best Local Similarity 26.4% Pred No. 0.035;
Matches 60; Conservative 28; Mismatches 83; Indels 56; Gaps 11;
Qy 33 PFGYKPGSKESIENLKDVNIWLIENRSDNLTG---GVRROGLDNPINNPGFCNYK 89
Db 33 PANHRSGSIED-----VEHIVVLMQENRSDHYFTGLRGVRFQDPHPV----- 76
Qy 90 NASDPSSGKYCTOAKDYDSVENDPDHVTGNLNLEFYTYTPNNGAIGASGVVADQSGFLN 149
Db 77 -----RLDDGRSVW---HQRKGDGTEVL-PFHPEADDLGMQFLEGLPHGWS 119
Qy 150 AQ--LND--YPKLAPPEATROVMGYTTEEVPTLVLDVDEFTTFNSWFSVCPGPTNPRL 205
Db 120 GQDAYHDAKYDRWLPAGKT-TTMAIYLTREDIPFHYALADTFVCDAYHCSFTGSTDPNRY 178
Qy 206 CALAGTAAGHGKND-----DDFLNYGISKSIFFAANEKGVSW 243
Db 179 YIW----SGHTGNDGAGGGPVLGNDGLGYDWTTP--ERLEAAGVSW 219

Search completed: December 4, 2001, 15:08:08
Job time: 68 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:07:00 ; Search time 13.39 Seconds
(without alignments)
1270.536 Million cell updates/sec

Title: US-09-426-072-2

Perfect score: 2501
Sequence: 1 MKSTALLTGLLASGLAS.....GTTLRDAPALKTPHTFSV 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	324.5	13.0	517	1	PHLC_MYCTU
2	318	12.7	700	1	PHLN_BURPS
3	308	12.3	514	1	PHLD_MYCTU
4	288	11.5	521	1	PHLB_MYCTU
5	287.5	11.5	730	1	PHLC_PSEAE
6	286	11.4	520	1	PHLA_MYCTU
7	264	10.6	692	1	PHLN_PSEAE
8	167.5	6.7	421	1	PHOX_KLULA
9	119.5	4.8	417	1	PHOA_PSPNG
10	116	4.6	412	1	PHOA_PENCH
11	112.5	4.5	488	1	PHB_ALCFA
12	111	4.4	428	1	GLNC_BRANA
13	110.5	4.4	729	1	COAL_PAVP9
14	109.5	4.4	729	1	COAL_PAVPK
15	107	4.3	1257	1	CCAA_BACTU
16	106.5	4.3	1341	1	YL78_YEAST
17	105.5	4.2	1211	1	ATHI_YEAST
18	105.5	4.2	4590	1	FATH_HUMAN
19	105	4.2	749	1	COA2_PAVPN
20	105	4.2	939	1	HEX_ADEP3
21	104.5	4.2	625	1	BGAL_LACSK
22	104.5	4.2	729	1	COAL_PAVPN
23	104.5	4.2	991	1	CPGL_PORGI
24	104.5	4.2	1122	1	ADPL_MYCGA
25	101.5	4.1	800	1	INLA_LISMO
26	101.5	4.1	1203	1	ALAB_ARATH
27	101	4.0	1146	1	KTXA_KLULA
28	100.5	4.0	736	1	CPG2_PORGI
29	99	4.0	434	1	GLN2_HORVU
30	99	4.0	1645	1	OMP_B_RICTY
31	98	3.9	430	1	GN2_PEA
32	97.5	3.9	458	1	YM51_YEAST
33	97	3.9	558	1	PAHA_CAEEL

34	97	3.9	584	1	D29B_ARATH
35	97	3.9	600	1	L765_ARATH
36	96.5	3.9	775	1	VP4_ROTFL
37	96.5	3.9	1006	1	RAT1_YEAST
38	96.5	3.9	2148	1	VIT1_AEDAE
39	96.5	3.9	4447	1	PXSK_BACSU
40	96	3.8	532	1	GSI_NEUCR
41	96	3.8	659	1	AMIA_STRPN
42	96	3.8	681	1	YADD_SCHPO
43	96	3.8	2329	1	YS89_CAEEL
44	95.5	3.8	976	1	AMY_BUTFI
45	95.5	3.8	1783	1	Y468_MYCGE

ALIGNMENTS

RESULT	1
PHLC_MYCTU	
ID	PHLC_MYCTU
AC	P95245
DT	20-AUG-2001 (Rel. 40, Created)
DT	20-AUG-2001 (Rel. 40, Last sequence update)
DE	PHOSPHOLIPASE C 3 PRECURSOR (EC 3.1.4.3).
GN	PLCC OR RV2349C OR MT2414 OR MTCY98.18C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1773;
ON	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	complete genome sequence.,"
RN	Nature 393:537-544(1998).
RP	[2]
RC	SEQUENCE FROM N.A.
RA	STRAIN=CDC 1551 / Oshkosh;
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA	Bisai W.;
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and
RL	laboratory strains.,"
CC	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC	-1- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = 1,2-
CC	DIACYLGLYCEROL + CHOLINE PHOSPHATE.
CC	-1- SIMILARITY: BELONGS TO THE BACTERIAL PHOSPHOLIPASE C FAMILY.
CC	-1- MISCELLANEOUS: POLYMORPHISM WAS DISCOVERED IN THE PHOSPHOLIPASE
CC	PLCA/B/C REGION.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to licensel@isb-sib.ch).
CC	-----
DR	EMBL; 283860; CAB06146.1; ALT_INIT.
DR	EMBL; AE007081; AAK46707.1; ALT_INIT.
DR	TIGR; MT2414; -.

Db 308 YTSWLDALTANPEWSKALLVDFDNDGFFDHVAPPAAPSLNKDGLRGKTKTADATLE 367
 QY 366 --AKNQKYLHFDLGRMPTWVISPYSKGYEQYGTDPVTKRPPAPYSATSVLKTL 421
 Db 368 WHTKGDIRNQPYGLGRVPMYVISPMSKGGWNS-----QVFDHTSVIRFL 415

RESULT 8

ID PHOX_KLULA STANDARD; PRT: 421 AA.
 AC P08540;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 DE POTENTIAL ACID PHOSPHATASE (EC 3.1.3.2).
 OS Kluyveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHS 2359 / IFO 1267 / NRRL Y-1140;
 RX MEDLINE=890341156; PubMed=3053697;
 RA Chang Y.-D., Dickson R.C.;
 RT "Primary structure of the lactose permease gene from the yeast
 Kluyveromyces lactis. Presence of an unusual transcript structure.";
 RL J. Biol. Chem. 263:16696-16703(1988).
 CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN
 CC ALCOHOL + ORTHOPHOSPHATE.
 CC -!- SIMILARITY: VERY STRONG, TO ASPERGILLUS NIGER AND PENICILLIUM
 CC CHRYSOGENUM PHOA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X06997; CAA30054.1; ALT_INIT.
 DR PIR; B31776; B31776.
 KW Hypothetical protein; Hydrolase.
 FT ACT_SITE 228 228 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 421 AA; 47193 MW; 2F7E614C943824B6 CRC64;

Query Match 6.78; Score 167.5; DB 1; Length 421;

Best Local Similarity 21.0%; Pred. No. 7.5e-05;
 Matches 89; Conservative 65; Mismatches 147; Indels 123; Gaps 23;
 QY 122 LEFGTGTYPNNGAIAAS-----GKVVAQDSG-----FLNAQLNDPKLAPETATROV 167
 Db 31 LRTVSTSPSLSEIASASATEVAEVSDVEGAERFFIFLENTDYDKAAGDES---- 86
 QY 168 MCYITEEVEPTLVLDVDEFTTFNSWFCVPGTNPRLCALAGTAGHKGNDLFLNGI 227
 Db 87 LSWLAEQGI-----TUTNYW--ALTHPSEPNYLASVGGDY--FALDDRFISMP 132
 QY 228 SSKSIFAANKEGVSMNYD-----GTNG-----EPEPDSLFFTYVNAQT 266
 Db 133 NVSNIVDLLOTKGISWAEQBSHPYAGFCQMFSNQTQVSDYVRKHNPLILFDNVNND 192
 QY 267 SR-SNVVPEVFFQDAYGLVLPKFTSYINPSCCTNTNSMHTGNVSGEVFYKQIYDAIR 325
 Db 193 TRLANIKNFEDFNNDYENKLPQVAFITPNM-----TNDGHDT-TIQFAGKWSKDFIAPLL 247
 QY 326 QCPQW-DKTLFLFTYDET-----GG-----FYDHVP----- 350
 Db 248 ENDYFEMEDTLVLITFDENETYGKKNVFSILLGGVIPDELAKTKDDTFYDHSQSALVEA 307
 QY 351 ----PPLAVRPDNLITETAKNGQKYYT-LHFDRLGGRMPTWVISPYSKGYEQYGTDPV 405

Db 308 NWDPLHLGHGDGDANVLIEIVANATNITNVEVD-----TTIMINE-TYIGLYNDYNE-- 358
 QY 406 TGKPAP-----SATSVLTKLYLWDIEDFTPRVAHSPSFDHLIGTLTREDAPIALKTP 459
 Db 359 --LPAFNTAINNRNQPILOSIKETWE-DEYSKQVSES-----YTTSTTTTTSADV--DA 409
 QY 460 HTFS 463
 Db 410 ETFS 413

RESULT 9

ID PHOA_ASPNG STANDARD; PRT: 417 AA.
 AC P34724;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).
 OS PHOA.
 GN Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=NRRL 3135 / VAN TIEGHEM / FICUUM;
 RX MEDLINE=95032134; PubMed=7945393;
 RA Ehrlich K.C., Montalbano B.G., Mullaney E.J., Dischinger H.C. Jr.,
 RA Ullah A.H.J.;
 RT "An acid phosphatase from Aspergillus ficuum has homology to
 Penicillium chrysogenum Phoa.";
 RL Biochem. Biophys. Res. Commun. 204:63-68(1994).
 CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN
 CC ALCOHOL + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: TO P.CHRYSOGENUM PHOA AND TO A KLUYVEROMYCES LACTIS
 CC ORF (AC P08540).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L20566; AAA62393.1; --
 DR Hydrolase; Glycoprotein; Signal.
 KW SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 417 ACID PHOSPHATASE.
 FT ACT_SITE 215 215 PROTON DONOR (BY SIMILARITY).
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 417 AA; 46735 MW; 896FD2AF541C69C8 CRC64;

Query Match 4.8%; Score 119.5; DB 1; Length 417;

Best Local Similarity 20.6%; Pred. No. 0.2;
 Matches 74; Conservative 43; Mismatches 137; Indels 105; Gaps 15;
 QY 155 YPKLAPETATROVNGYIT-----EEEYPTLVD---LV 183
 Db 26 YPSLAETRAAQATVQPVSPVSNVKGTLFNRFNWILENTDEDAATAATEHLPLVAKKGLLL 85
 QY 184 DFEFTTFSWFCVPGTNPRLCALAGTAGHKGNDLFLNGIISKSIFEAAANEKGVSW 243

Db 86 NNF-----W--AVTHPSPN-YCRHPLGLDTEGMDNDHFDHQPSPNVTIADLFTKNIW 137
 QY 244 -----LNYDGTNGEFPDS-----LFFTVNOTSR-SNVVPEVFOQAYL 283
 Db 138 GEYCEGLPYPGYGYRPEGANDYVRNRNPLILFSDVEDALRLQIKNESFVDDLEN 197
 QY 284 GVLKFSYINPSCCTNTNSMHPGNSVGEVFKQIYDAIROGPQWDK-TLLFTTDET 342
 Db 198 HRLPOYMFITPNM-----TNDGHD-TNITSGDWTFGLSELLENDYFTKTLMLTDET 252
 QY 343 GGFY-----DHVPPPLAVRPDNLTYT-----ETAKNGQYTLHFDRLGGRMPT 385
 Db 253 GYIEIGNNIVTLLGAVPDDLLGTDDTFYTHYSVIASLSTNWGLPSLGRWDCGANLFS 312
 QY 386 WVI-----SPYSKKGYIEQYGTDPVTKKAPYSA-TSVLTKL 421
 Db 313 WLAKKTYGVNVEVDSNLYMETHWGLFSLDDDDYSEYAGWPVPTTDASCAGNGLSTV 371

RESULT 10
 ID PHOA_PENCH STANDARD; PRT; 412 AA.
 AC P37274;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).
 OS PHOA.
 GN Penicillium chrysogenum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
 OX NCBI_TaxID=5076;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=9225342; PubMed=1563629;
 RA Haas H., Redl B., Friedlin E., Stoeffler G.;
 RT "Isolation and analysis of the Penicillium chrysogenum phoA gene
 encoding a secreted phosphate-repressible acid phosphatase.";
 RL Gene 113.129-133(1992).
 CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN
 CC ALCOHOL + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: TO A NIGER PHOA AND TO A KLUYVEROMYCES LACTIS ORF
 CC (AC P08540).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80366; AAA33693.1; -
 DR FIRM, JN0319, JN0319.
 KW Hydrolase; Glycoprotein; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 412 PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE.
 FT ACT_SITE 214 214 PROTON DONOR (BY SIMILARITY).
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 412 AA; 45628 MW; 5659AF07BDDFD9 CRC64;

Query Match 4.6%; Score 116; DB 1; Length 412;
 Best Local Similarity 19.9%; Pred. No. 0.35;
 Matches 67; Conservative 40; Mismatches 111; Indels 118; Gaps 15;

QY 179 LVDLVDEFTTNSWFSVCPGPTNPNRLCALAGTAAGHKNDLFLNYGISKSIFFAANE 238
 Db 75 LSKLAKEGILLTNVFA-ISHPSQPN-YCASAG-GDTFGMDNDLFLQIPSNVSTIADLFT 131
 QY 239 KGVSW-----LNYDGTNGEFPDSLFTTYV-----NOTSRNVVP--VEN---FF 278
 Db 132 KHSWGEQEDMPAGYQGRYPLSGNQYRKHNPLVLENSVTDADVPRQIKNFTFY 191
 QY 279 QDAYLGVLPKFSYINPSCCTNTNSMHPGNSVGEVFKQIYDAIROGPQWDK-TLLFI 337
 Db 192 DDLKHSLSLFOHMFITPNM-----TNDADHT-NITVAGNWDVRFSLPKKNEFTKDSVL 246
 QY 338 TYDETGGVDH-----VPPPLAVRPDNLTYTETAKNGQYTLHFDRLGGRMPT 385
 Db 247 TFDG-GDYSYPNRVFSLVGGAIPEHLKGTDDTFY-----HYSIVASISAN 294
 QY 386 WVI-----SPYSKKGYIEQYGTDPVT 406
 Db 295 WGLPSLGRWDCGANLLKMWADKTGYVNWEDTSNVYLNETYPGPMSTDNYSKNAVPATK 354
 QY 407 GK-----PAPYSATS 416
 Db 355 GKCSAGHGTAEVKNYTHGLQPTYDASVPYDVTS 390
 RESULT 11
 ID PHB_ALCFA STANDARD; PRT; 488 AA.
 AC P12625;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE POLY(3-HYDROXYBUTYRATE) DEPOLYMERASE PRECURSOR (EC 3.-.-.-) (PHB
 DE DEPOLYMERASE).
 OS Alcaligenes faecalis.
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Alcaligenes.
 OX NCBI_TaxID=511;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TI;
 RX MEDLINE=89123016; PubMed=2644188;
 RA Saito T., Suzuki K., Yamamoto J., Fukui T., Miwa K., Tomita K.,
 RA Nakanishi S., Odani S., Suzuki J.-I., Ishikawa K.;
 RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
 the gene for poly(3-hydroxybutyrate) depolymerase from Alcaligenes
 faecalis.";
 RT J. Bacteriol. 171:184-189(1989).
 RN [2]
 RP FIBRONECTIN TYPE III DOMAIN.
 RX MEDLINE=93028390; PubMed=1409594;
 RA Bork P., Doolittle R.F.;
 RT "Proposed acquisition of an animal protein domain by bacteria.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8990-8994(1992).
 CC -!- FUNCTION: THIS PROTEIN DEGRADES WATER-INSOLUBLE AND WATER-SOLUBLE
 CC PHB TO MONOMERIC D(-)-3-HYDROXYBUTYRATE.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J04223; AAA21974.1; -
 DR FIRM; A32235; A32235.
 HSSP; P02751; 1TTG.

QY 315 VFVKQIVDAIROGP-QWDKTLITFYDETGGEYDHPVPLAVRPDNLTYTETAKNGO-KY 372
DB 477 A-----TAIRPAQGVYNTPMNFESNGGPELPPIVPTA-----DTQYNDDEPNGAIRF 525
QY 373 TLHPDRLGGRMT-----WISPSYKGGYI--EQYG-----TPVPT 406
DB 526 TMGYOH--GQLTSSQELEYTFNPSQKGRAPKQOFNOQAPLNLENTNNGTLLPSDPIG 583
QY 407 GKP-----APYSA---TSVLKTLGCLWLDIE---DETPRV 434
DB 584 GKPMMHFMNTLNTYGLPLANTAPFPNGQIWDKELDTLKLPL 628

RESULT 15

CCAA_BACTU STANDARD; PRT; 1257 AA.
AC Q45754;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PESTICIDIAL CRYSTAL PROTEIN CRY12AA (INSECTICIDAL DELTA-ENDOTOXIN
DE CRY11A(A)) (CRYSTALLINE ENTOMOCIDAL PROTOXIN) (142 KDA CRYSTAL
DE PROTEIN).
GN CRY12AA OR CRY11A(A) OR CRYVB.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.
RC STRAIN-NRRL B-18244 / PS33f2;
RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,
RA Sick A.J.;
RT "Novel Bacillus thuringiensis microbes active against nematodes, and
RT genes encoding novel nematode-active toxins cloned from Bacillus
RT thuringi."
RL Patent number EP0462721, 27-DEC-1991.
CC -!- FUNCTION: ENDOTOXIN WITH NEMATICIDAL ACTIVITY.
CC -!- DEVELOPMENTAL STAGE: THE CRYSTALL PROTEIN IS PRODUCED DURING
CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; L07027; AAA22355.1; ..
DR InterPro; IPR001178; Endotoxin.
DR Pfam; PF00555; endotoxin; 2.
KW Toxin; Sporulation.
SQ SEQUENCE 1257 AA; 142265 MW; 3D988FFC6C0E3981 CRC64;

Query Match 4.3%; Score 107; DB 1; Length 1257;
Best Local Similarity 19.3%; Pred. No. 6.8; 166; Indels 166; Gaps 23;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;
QY 6 LITGLGLASLG-LASPTVSEYTSVREAPFGYKPGSKESIENLXDKVENIVWLLEN--- 61
DB 74 VLSVLTLPVGTVAASASTVSVFWPKIFGDKPNKAKNIFEELKPKQIEALIQODITNQD 133
QY 62 ----RSDNLLGVRRQGLDNFPIINNGPCFNKASDPSSGKYCTQAKDYDSVND----- 112
DB 134 AINOKKFDLSQKTLINLYTV--AIDNDYVTAK-----TOLENLSILTSISIF 180

QY 113 -PDHSVTGNLLEFYGTYPNNGAIAAGKYV-ADQSGFLNAQLNDYKPLAPEATROVNGY 170
DB 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234
QY 171 YTEEEVPTLVLDVDEFTTF--NSWFSCVPGPTNPNRLCALAGTAAGHGKNDDEFLNYGIS 228
DB 235 HTEAVIKAFENGLDKFKSLDWNST-----NKRANY--- 264
QY 229 SKSIFEAANEKGVSWLNYDGTNGEFPDPSLFFTYVN-----QTSRNVVVPVENF 277
DB 265 IKGTEKVLVDLVALMPTFDPDHQKEVEIEFTRTISSPIQPVPKNMONTSSSIIVPSDLF 324
QY 278 FQDAYLGLPK--ES-----YINPSCGT-----NTNSMHPGN 309
DB 325 H---YQGDVLKLEFSTRDNDGLAKFTGIRNTFFKSPNTHYHYHDFSYNTQS---SGN 378
QY 310 VSYG-----EVFVKQIYD-----AIRQGPQWDKTLIF 336
DB 379 ISRGSSNPIDLNPIISTCIRNSFYKAAGSSVLVNFKDGTOGYAFAQAP----- 430
QY 337 IYDETGGEYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLHEDRLGGRMPTWVI 388
DB 431 -----TGGAWDHSFIESDCAPEGHKLNYYTSPGDTLDRDFINVTLLISTPTINELSTEKI 485
QY 389 SPY-SKKGYIEQYGDPTVTGKP 409
DB 486 KGFPAPKGYIKNOGIMKYYGKP 507

Search completed: December 4, 2001, 15:08:21
Job time: 81 sec

Qy	51	VENTVWLLLENRSFNIILGVR--ROGLDNPIINGPCNWKYKNASDPGSKCYCTQAKDYDS	108
Db	14	IKTTVLQVQENSDPHTLGMFKELNREIDYQTKDPKSNVSSSDTNSLR--VVFQDSQ	71
Qy	109	VFN-DPDHVSVTGNLLEFY-----GTYTPNNGAISGKVVQASQGLF-NAQLNDYPLKAP	160
Db	72	YVNPDPCHSIQDIYEQVFGKPWDGKPDNPCH-----PNMSGFAQNAERN-----K	118
Qy	161	EEATPROVMGYTTEEBEPTLVLDVDETFITNSKWFSCVGPPIPNRLCALAGTAAGHKND	220
Db	119	KGSSAVMNGFPNALPVYKELVQNFACDHRFASVSPASTQPNRLYVHSATSHGATSNOK	178
Qy	221	DFLNGLYSKSTIPAEANEKGVSWLNYDGTNGEFPDLSLP-----TYVNGTSSSNVVPY	274
Db	179	KLLEGGFPQKTFISLDEAGFSF---GIYYQFPSTFLFNLRLKLYLTHFYQGI---	231
Qy	275	ENFPQAYLGLVPKFSYNP-----SCCGTNTNSMPTGNSYGEVFKQIYDAIRQSPD	331
Db	232	-QFKKDCKEGLPNYVVVEQRFDFLLSTPANDHPSHDVSEGOKLVKEVYALRSSPQPN	290
Qy	332	KTLFLITDETGGFDVHPPLPAVRPDNLVITETAKNGQRYTLHFDRLGGRMPTWISPY	391

Db	291	ELIFITYDEHGFFDVHVPV-----DGVNPDGILGPPYFNFENLUGVRVPTFFISW	348
QY	392	SKGYIEQTGDPVTGKPAPE-----YSATSVLKTGLYMDIEDF-TPRVAHSPFDHLI-	444
Db	347	IEPTGV-----THGNGPYPRSQYEHSSIPATVKTIFKLDFLSKRDWSWAGTFESVIT	399
QY	445	GTTLRDPAIAIKTP	459
Db	400	RDSPROCPETLSTP	414
RESULT	3		
Q9JLM5		PRELIMINARY;	PRT; 533 AA.
Q9JLM5			
DC	01-OCT-2000	(TRENBLrel. 15, Created)	
AT	01-OCT-2000	(TRENBLrel. 15, Last sequence update)	
DT	01-OCT-2000	(TRENBLrel. 15, Last annotation update)	
DE	F10K1.5	PROTEIN.	
GN	F10K1.5		
OS	Arabidopsis thaliana	(Mouse-ear cress).	
OC	Eukaryota;	Viridiplantae;	Streptophyta;
OC	Spermatophyta;	Magnoliophyta;	eudicotyledons;
OC	eurosid II;	Brassicales;	Brassicaceae;
OC	NCBI_TaxID=3702;		
RC	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Liu S.X.,	Chan A.,	Yu G.,
RA	Sakano H.,	Toriiumi M.,	Chung M.,
RA	Vaysberg M.,	Altafi H.,	Brooks S.,
RA	Conway A.,	Hansen N.,	Johnson-Hopson C.,
RA	Miranda M.,	Nguyen M.,	Palm C.J.,
RA	Ecker J.R.,	Pederspiel N.A.,	Theologis A.,
RT	"The sequence of BAC F10K1	from Arabidopsis	thaliana chromosome 1."
RT	Submitted (APR-2000)	to the EMBL/GenBank/DBJ	databases.
RL	[2]		
RL	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RC	STRAIN=CV. COLUMBIA;		
RL	Theologis A.;		
RL	Submitted (APR-2000)	to the EMBL/GenBank/DBJ	databases.
RL	[3]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RC	STRAIN=CV. COLUMBIA;		
RL	Theologis A.		
RL	Submitted (JUN-2000)	to the EMBL/GenBank/DBJ	databases.
RL	[4]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RC	STRAIN=CV. COLUMBIA;		
RL	Theologis A.		
RL	Submitted (JUL-2000)	to the EMBL/GenBank/DBJ	databases.
DR	EMBL; AC067971,	AAF82197.1;	
SQ	SEQUENCE	533 AA; 60038 MW; 6D7ADA226E2E5215	CR664;
Query Match	18.1%;	Score 453.5;	DB 10; Length 533;
Best Local Similarity	30.9%;	Pred. No. 8e-27;	
Matches	139;	Conservative	77; Mismatches 167; Indels 67; Gaps
QY	34	FGYKPGKSEIE-----NLKDKVENTVLLLENRSFDNILGVR--RGLDNPINNG	83
Db	13	FYLLIASSQSIFFKNSOKPIKIQGTPITVVYVNNRSFDHILKSTREIDG--LTG	70
QY	84	PFCNKNASDPSGGKYCTQAKDYDF--NPDHSVNTGNMLFYGYTPNNGATASGVV	141
Db	71	KESNPLNVSDPSKKIFVSD--DAVFVMDPGHSCQAIREQIFG----SNDTSGDPKM-	122
QY	142	ADQSGFLNQLNDYFKLAPEATQVMGYIIEEVPTLVLDVDEFTFNFSWFCVPGPTN	201
Db	123	---NGFAQGSSEWEPGMA-----KNVMSGFPEVLPIVYTELANEFGVDFRWFA	174
QY	202	PNRLCALAGTAGHGKNDDDFLNYIGSSKSIIEAANEKGVSWLNTDGTNGEFPD	261

```
Db 175 PNFYVHSATSHGCSNVKDLVKGFPQKTFIDSLDENGSLF-----GIYYONIPATFFFK 230
Qy 262 YNQTNRSNVVPVEN-----FFQDAYLGVLPKFSYINPSCCGTN---TNSMHPGTVSYGE 314
Db 231 SLRRL--KHLVKFHSYALKFKDAKGLKLPNYSVVEQRYFDIDLFPANDHPSHDVAAGQ 288
Qy 315 VFVKQIYDAIRGQPMQDKTLITFYDETCGFYDHVPPPL--AVRPDNLTYTETAKNGQY 372
Db 289 REVKEVETLRSSPQWKEMALLITYDEHGGFYDHVPTPVKGVNPDGII-----GPDPE 342
Qy 373 TLHFDRLGRMPTWVISPYSKGYI--EQYGTDPVTKGPKAPYSA---TSVLKTLGYLWDI 427
Db 343 YFGDRLGVVRVPTFLISSWIEKGTVIHEPEG-----PPHSQFHSIPATVKLFLNL 395
Qy 428 ED--FTPRVHSPSFDHL--IGTTLREDAP 453
Db 396 KSHFLTKRDWAGTFEKYFRIRDSPRQDCP 425

RESULT 4
Q9S816 PRELIMINARY; PRT; 521 AA.
AC Q9S816;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE PUTATIVE PHOSPHOLIPASE.
GN T12J13.18 OR T2J1P5.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Rowning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence.";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009327; AAF03477.1; -.
DR EMBL; AC009895; AAF01581.1; -.
SQ SEQUENCE 521 AA; 59071 MW; 1A39ACEF6807B8EF CRC64;

Query Match 17.58; Score 437; DB 10; Length 521;
Best Local Similarity 30.08; Pred. No. 1.5e-25;
Matches 133; Conservative 73; Mismatches 161; Indels 76; Gaps 18;

Qy 51 VENTVWLILNRSFONILG-----GVRRO-----GLDNPIN--NGPFCNYKNAS 92
Db 13 IKTIVLVQENSFDHTLIGWFKELNREIDGVKMSDKPFGFSSDLNHNHVVFGDSQYV 72
Qy 93 DPSSGKVCYQAKD--DSVFNDDPHSVTGNLLEFYCTYTPNNGATASGVVADQSGFL-NA 150
Db 73 DPNPGH---SIRDIIEQYEGKPDWS-----GHPDNPGE-----ATWSGFAQNA 113
Qy 151 QLNDYPKLAPEATRVQVMGYTTEEVPTLVLDVDEFTTNSWFSVCPGPTNPNRICALAG 210
Db 114 ERK-----MKGSSAVMNGFKPDALPVYKELVQNFACIDRFASVPGATQPNRLFIHSA 167
Qy 211 TAAGHKNDLNGYISGSKSIFEANEKGVSNLNYDGTNGEREPSLFTYVNO---TS 267
Db 168 TSHGTTNKRKLLIEGFPQKTFIESLDEAGTFI---GIYQCEPTLYRNLKLYLT 223
Qy 268 RSNVVPVENFQDAYLGVLPKFSYI-----NPSCCGTNTNSMHPGTVSYGEVFK 318
Db 224 RFHDYGLQ-FKKDCKEGLNPNYVVVEQRYDYDLLNPA-----NDDHPSHDVSQGLVK 276
Qy 319 QIYDAIRGQPMQDKTLITFYDETCGFYDHVPPPLAVRPDNLTYTETAKNGQYTLHFD 378
Db 277 EYVEALRSSPQWKEMALLITYDEHGGFYDHVPTPL-----DGVNPDGILGPPPYNFEFR 332
```

```
Qy 379 LGRMPTWVISPYSKGYIEQYGTDPVTKGPKAPYSATSVLTKTLGYLWDIEDF--TPRVAHS 437
Db 333 LGRVPTTFISPIWIEPCTVLHGNGPYL--MSQYEHSSIPATVKIKFLKDFLTKRDSWA 390
Qy 438 PSFDHLI--GTTLRDAPIALKTP 459
Db 391 GTFESVITRNSPRQDCPETLSNP 413

RESULT 5
Q9SMN8 PRELIMINARY; PRT; 462 AA.
AC Q9SMN8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE HYPOTHETICAL 51.3 KDA PROTEIN.
GN T8P19.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choise N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133315; CAB62350.1; -.
KW Hypothetical protein.
SQ SEQUENCE 462 AA; 51284 MW; EE6492B529991095 CRC64;
```

```
Query Match 17.48; Score 436; DB 10; Length 462;
Best Local Similarity 32.08; Pred. No. 1.5e-25;
Matches 125; Conservative 61; Mismatches 165; Indels 40; Gaps 13;

Qy 78 NPINN---GFCNYKNASDPSSGSKYCTQAKDYDSVFNDDPHSVTGNLLEFYCTYTPNNGA 134
Db 6 NPTINGVTGECN---PVPNSTQTICFTSDAEFVDPDPGHSFEAVEQVFGS---GPGQ 58
Qy 135 IASGVVADQSGFLNANLNDYKLAPEATRVQVMGYTTEEVPTLVLDVDEFTTNSWFS 194
Db 59 IPS-----MKGFEQALIS-----MFCNLSEYVMKGFPEAVPYAELVKEAFEDRWFS 107
Qy 195 CVPGPTNPNRICALAGTAAGHKNDLNGYISGSKSIFEANEKGVSNLNYDGTNGEFE 254
Db 108 SIPGPTQPNRLVYVATSGSTSHVKKQLAOGYPQKTFIDSLHNSNIDF----GIYFONI 163
Qy 255 PDSLEFFTYNQTSR--SNVVPVE--NFODAYLGVLPKFSYINP---SCCGTNTNSMHPGTV 309
Db 164 PTLTLYRNLRLQKLYIFNLHQLKFKKDAKGLPLSLVIEPRYFDLGLPANDDHPSHD 223
Qy 310 VSYGEVFPVQIYDAIRGQPMQDKTLITFYDETCGFYDHVPPPLAVRPDNLTYTETAKNG 369
Db 224 VANGQKLKVEYVEALRSSPQWKEMALLITYDEHGGFYDHVPTPVKGVNPDGNTGAPGF 283
Qy 370 QKTYLHFDRLGRMPTWVISPYSKGYIEQYGTDPVTKGPKAPYSATSVLTKTLGYLWDIED 429
Db 284 FK----FDRLGVRVPTWVSPWIKGTVYSEAKGPT--ESSEYEHSSIPATIKKFLNLS 337
Qy 430 --FTPRVHSPSFDHLIG--TTLRDAPIAL 456
Db 338 NFLTHRDWAATFEDVWSHLTTPRTDCPMTL 368

RESULT 6
```


OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
 OC Burkholderia.
 OX NCBI_TaxID=28450;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPT1;
 RX MEDLINE=99454884; PubMed=10523590;
 RA Korbsrisate S., Suwanasai N., Leelaporn A., Ezaki T., Kawamura Y.,
 RA Sarasombath S.;
 RT "Cloning and characterization of a nonhemolytic phospholipase C gene
 from Burkholderia pseudomallei";
 RL J. Clin. Microbiol. 37:3742-3745(1999).
 DR EMBL: AF107252; AAF17299.1; ..
 SQ SEQUENCE 700 AA; 77190 MW; 5DAE6FCD0F129B CRC64;

Query Match 12.7%; Score 318; DB 2; Length 700;
 Best Local Similarity 27.8%; Pred. No. 3.5e-16;
 Matches 126; Conservative 55; Mismatches 160; Indels 112; Gaps 24;

QY 6 LLTGLGLASGLASPTSEYTSVREA---PFGYKPGSKESIEHLKDKVENIVLWILENR 62
 DB 12 LAAGTGAALQLFPV-----IREALIPANRTGT-----IRD-VEHIVILMOENR 58
 QY 63 SFONILG---GVRROGLDNPI---NNGPFCNKNASDPSSGKYCTQAKDYDSVFNPDHS 116
 DB 59 SFDHYFCKLRGVRGDPRLTLQNGKPVHFQVLLGPA-----ELLPPHPDAS 107
 QY 117 VTGNLNF-----YCTTPNNGAIAAGKVVADQSGFLNAQLNDYKLAPEEATROVMGY 171
 DB 108 NLG--WQFLQDLPHG--WQTHGAWNKR-----YDRWIANKGT--TTMAYL 148
 QY 172 TEEVEPTLVLDVDEFTFNSFCVPGTPNPNRLCALAGTAGAGHKNDDEFL-----NYG 226
 DB 149 ERGDIPHYQLADAFICDAVHCSIPSSDPNRYVMYWGVDGAGGPGVNLGNEAGY 208
 QY 227 ISSKSIPEANKEGVSMLND--GTN-----GFEFDSLFF-----260
 DB 209 WSTVP--ETLEGAGVSWKIVODIGTGLDAAGSWGWTQNPYIGNYGDNALLXFQVYRNAQ 266
 QY 261 ---TYNQRSNRYVVENEF---QDAYLGLVPKFSYINPSCCGTNTNSMHTGNYSG 313
 DB 267 GSPLYDKARTGNVAGGTLFDVLQDDVKNGTLPQVSWI-----CAPEAYSEHPNWPAN 322
 QY 314 EVFKQIYDAIRGGPQ--WDKTLFTYDETFGFDYHVPPLA--VRPDNLTYETAKNGQ 370
 DB 323 ANIVEQVLKALTSNPDVWSKTLFTYDENOGFDFHVPAPQSRDGLSTVSTA--GE 380
 QY 371 KY---TLHFD---RLGGRMPTWISPYSKGYI 397
 DB 381 IFPGDAAHMAGVGLGPRVPMVYVSPWTKGGWV 413

RESULT 9

Q9XB13
 ID Q9XB13 PRELIMINARY; PRT; 514 AA.
 AC Q9XB13;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE PHOSPHOLIPASE.
 GN RV2D-PLC.
 OS Mycobacterium bovis BCG.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=33892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BCG-PASTEUR;
 RX MEDLINE=99255698; PubMed=10320585;
 RA Gordon S.V., Brosch R., Billault A., Garnier T., Eiglmeier K.,
 RA Cole S.T.;
 RT "Identification of variable regions in the genomes of tubercle bacilli

RT using bacterial artificial chromosome arrays.";
 RL Mol. Microbiol. 32:643-655(1999).
 DR EMBL: Y18606; CAB44656.1; ..
 SQ SEQUENCE 514 AA; 55770 MW; 9D99A466D312BA30 CRC64;

Query Match 12.3%; Score 308; DB 2; Length 514;
 Best Local Similarity 28.0%; Pred. No. 1.3e-15;
 Matches 124; Conservative 45; Mismatches 156; Indels 118; Gaps 20;

QY 9 GLGLASGLASPTSEYTSVREAPRGYKPGSKESIEHLKDKVENIVLWILENRSFONIL 68
 DB 22 GAGALMS--FAGPVI-----EKAYGAGPCS---GHLD-IEHFVFFMQENRSFDHYF 67
 QY 69 GGVRRGLDNPIINGPCNKNASDPSSGKYCTQAKDYDSV-----F 110
 DB 68 GTL--SGTGFNTVSPLFQKQWNP-----MTQALDATGVTMPYRFTTRGPFIDGACV 119
 QY 111 NDPDHSVTGNLNFYGTTPNNGAIAAGKVVADQSGFLNAQLNDYKLAPEEATROVMGY 170
 DB 120 NDPDHS-----WYAMHESWNGV-----NDNLPAQ---AKTSAAHTPTVMGY 160
 QY 171 YTEEVEPTLVLDVDEFTFNSFCVPGTPNPNRLCALAGTAGAGHKNDDEFL-----223
 DB 161 YTRQDIPHYLLADAFVCDYFCYGLGPTLPNRLYMLSATIDPDQGGNGPELOSPTFQ 220
 QY 224 ---NYGISSKSIPEANKEGVSMLNDYDGTNGEPEPDSLFTYVNOTS-----267
 DB 221 VRFGWRIMPQNLSDA---GVSKYV--RNKTLGPISSVLTYGSLVTSFKQSDAPRSDL 274
 QY 268 -RSNVVP--VENFODAYLGLVPKFSYINPSCCGTNTNSMHTGNYSGEVFKQIYDAI 324
 DB 275 VREGVAPSPASFAADVLNRLPRVSWIPNV---LESEHPAPVAPAAAGAFVAILRIL 330
 QY 325 RQGPQ--WDKTLFTYDETFGFDYHVPPLA--VRPDNLTYETAKNGQVTL--HEDR---378
 DB 331 LANPAPWEKTLIVSDYDNGGFDHVPVATA-----PAGTPGEYTVDPIDQVPGS 381
 QY 379 -----LGRMPTWISPYSK 393
 DB 382 GGIRGPIGLGFRVPCFVISPSYR 404

RESULT 10

P95246
 ID P95246 PRELIMINARY; PRT; 512 AA.
 AC P95246;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE PLCB.
 GN PLCB OR RV2350C OR MTCY98.19C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 DR EMBL: Z83860; CAB06147.1; ..
 DR TubercuList; RV2350c; ..

Db 383 SGGIRGLP---GLGPRVPCIVISPSYR 406

RESULT 12

Q50561 PRELIMINARY; PRT; 521 AA.

AC Q50561;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DE PHOSPHOLIPASE C.

GN MPCB.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RA JOHNSON K.A.;

RA STRAIN=H37RV;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U49511; AAC18944.1; -

SQ SEQUENCE 521 AA; 56424 MW; 7C915196E6914256 CRC64;

Query Match 10.8%; Score 271; DB 2; Length 521;

Best Local Similarity 25.5%; Pred. No. 9.7e-13;

Matches 138; Conservative 60; Mismatches 186; Indels 158; Gaps 29;

QY 14 ASLGASPVSEYTSVRAPPGYKPGSKESIKDKVENIWLLENSPDNLIGVVR- 72

Db 27 AFMSLAGPIIKAYGA-----GCPG-----HLFD-IEHIVLLMOENRSEFDHFGTUSD 74

QY 73 RQGLD-----NPNNGPFCNTKNASDPSSGKCYTQAKYDSV-----F 110

Db 75 TRGDDTTPPVFAQSGWNPMT-----QAVDPAG-----VTLPRYRFTTRGPLVAGECV 123

QY 111 NDPHSVTGNLEFYGTYPNNGAIASGVVADQSGFLNAQLNDYKPLAPEATQVMGY 170

Db 124 NDPHSWIGMHNSW-----NGGA-----NDNMLPAQVPFSPLOGNVPVT---MGF 165

QY 171 YTEEVPTLVLDVDEFTFNFSWFCVPGPTNPNRLCALA-----GTAAGH---GKNDDD 221

Db 166 YTRDLPPIHYLLADFTVCDCYFCSLGGTTPNRLYWNAMWIDPDGDTGGPVLIEPNIQ 225

QY 222 FLNGYSSKSIIEAANEKGVSW-----LNYDCTGEEF---EPDSLFTFY 262

Db 226 LQHY--SWRIMPENLEDAGVSKVYQNKLLGALNNTVVYNGLVNDYKQAADPRS----- 278

QY 263 VNQTSRNVVPE--NFFQDAYLGVLKFSYINPSCCGTNTNSMHT--GNVSYGEVFKQI 320

Db 279 --NLARGISPTYPDLDAADVRNRLTKVSWLPGF-----LLSEHPAFVNVGAV---AI 329

QY 321 YDAIR---QGPQ-WDKTLFFITDEGGYDHPVP-----LAVRPDLNLTETAK 367

Db 330 VDARILLNSPAWSEKTAALVNDENGFFDHPVPPPGTGPGEFVTPD---IDSVPGS 387

QY 368 NGQYTLHFDRLGGRMPTWVISPYSKKGIEQYGTDPVT-----GKPAPYSATSV 417

Db 388 GGIRGAI---GLGPRVPCIFISPYT--GPLWHDFTDHTSLKILRAREGVVPNLTAWR 442

QY 418 LKTYLWLDIEF--TPRAVHSPFDH-----LIGTLREDAPTAKTP 459

Db 443 DATYGDITSTFNAAPPNPKPNLDHPRAQCALPKLQCVPNVGLTGTAKTAIPYRVFP 502

QY 460 HT 461

Db 503 QS 504

RESULT 13

Q9A412

ID

PRELIMINARY;

PRT; 686 AA.

RESULT 14

Q9L5C5

ID

Q9L5C5;

AC

PRELIMINARY;

PRT; 577 AA.

AC Q9A412;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DE PHOSPHOLIPASE C.

GN CC3031.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI_TaxID=69394;

RN [1]

RP SEQUENCE FROM N.A.

EX MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C.; Feldblyum T.V.; Laub M.T.; Paulsen I.T.; Nelson K.E.;

RA Eisen J.; Heidelberg J.F.; Alley M.R.K.; Ohta N.; Maddock J.R.;

RA Polocka I.; Nelson W.C.; Newton A.; Stephens C.; Phadke N.D.; Ely B.;

RA DeBoy R.T.; Dodson R.J.; Durkin A.S.; Gwinn M.L.; Haft D.H.;

RA Kolonay J.F.; Smit J.; Craven M.B.; Khouri H.; Shetty J.; Berry K.;

RA Utterback T.; Tran K.; Wolf A.; Vamathevan J.; Ermolaeva M.; White O.;

RA Salzberg S.L.; Venter J.C.; Shapiro L.; Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL proc. Natl Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL: A8005965; AAK24993.1; -

TI:GR; CC3031; -

KW Complete proteome.

SQ SEQUENCE 686 AA; 75591 MW; 00EF60FAA00F8AFF CRC64;

Query Match 10.5%; Score 263.5; DB 2; Length 686;

Best Local Similarity 24.8%; Pred. No. 5.5e-12;

Matches 129; Conservative 77; Mismatches 186; Indels 129; Gaps 26;

QY 5 ALITGGLIA---SIGLASPVSEYTSVRAPPGYKPGSKESIKDKVENIWLLEN 61

Db 8 SLLAALGALPALARAADAIDAVRS-----GTIQD-VEHVILMOEN 50

QY 62 RSPDNLG---GVRQGLDNPINNGPFCNKNASDPSSGKCYTQAKYDSVND--PDHISV 117

Db 51 RSPDHFVGLNGVRGFGDRFI-----PVRDAAGRESSVVFQVQMSKDKLLAPPLNSA 104

QY 118 -TGNLEFYGTYPNNGAIASGVVADQSGFLNAQLNDYKPLAPEATQVMGYTEEV 176

Db 105 QTFAMRVES--TPHSWTD-----QDAWQGRMDRW----PDAKKPWSMGYFQADI 151

QY 177 PTLVDLVDEFTFNFSWFCVPGPTNPNRLCALAGTAAGHGK-----NDDDFLNYGIS 228

Db 152 PFQALADAFITLADYHCTGTGTNTNRLFLWTGNDLGKAGGPGTSNSHDFAEKGA 211

QY 229 SKSIF-----EAANEKGVSWLNDYGTNGEGEPDLSL--FFTY-----VNQTSR 268

Db 212 KESYTWTTTPPERLLQAGVSWRIYQDMADNFTONPLAGFKAYRDAYKDLFGSPNRLKQLG 271

QY 269 SNVYVENEFOADYLGVLKFSYINPSCCGTNTNSMHT--GNVSYGEVFKQIYDAIRQG 327

Db 272 S-TWHLDKLREDVWNGRLPQVSWI-----TAPADSEHPNPPSSPAQAGADYARVIDALTAD 326

QY 328 PQ-WDKTLFFITDEGGYDHPVPPPLAVRPD-----NLTYYTET 365

Db 327 PKVWARTVFLVMFDEGDFDHPVPPAPPSPYSDASGKLLGASTVDLTAEHHRVNRNTEARS 386

QY 366 AKN---GQYTLHFDRLGGRMPTWVISPYSKKGIEQYGTDPVTGKPAPYSATSVLTKLG 422

Db 387 ERDLMGRPY----GLGPRVPLVISPWSRGVWNS-----QVFDHTSVLRELE 431

QY 423 YLWDIEDFTPRVAHSPFDHLIG--TTLREDAPIALKTPHT 461

Db 432 QRFQVME--PNI--SPWRRVAVCGDLTTCFD-----FKTPNT 463

Query Match 8.1%; Score 203.5; DB 2; Length 514;
Best Local Similarity 21.7%; Pred. No. 1.6e-07;
Matches 117; Conservative 68; Mismatches 156; Gaps
Indels 199;

8 LGIL-SLSIAT--TTFATDVNNS-----KPNNDYGTGLVKIEQKLFENNANTLKTTTPIKHVVI 60

[illegible]

Search completed: December 4, 2001, 15:09:19
Job time: 139 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:07:00 ; Search time 15.89 Seconds
(without alignments)
657.113 Million cell updates/sec

Title: US-09-426-072-2
Sequence: 1 MKSTALLTGLGLASGLAS.....GTTLRDAPIALKTPHTFSV 464

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2501	100.0	464	US-09-426-072-2	Sequence 2, Appl
2	111.5	4.5	1041	US-08-494-714-2	Sequence 2, Appl
3	111.5	4.5	1041	PCT-US96-10782-2	Sequence 2, Appl
4	109.5	4.4	579	5223424-13	Patent No. 5223424
5	107	4.3	1257	US-08-049-783-2	Sequence 2, Appl
6	107	4.3	1257	US-08-158-232-6	Sequence 6, Appl
7	107	4.3	1257	US-08-304-626-6	Sequence 6, Appl
8	107	4.3	1257	US-08-316-301A-6	Sequence 6, Appl
9	107	4.3	1257	US-08-611-928-6	Sequence 6, Appl
10	107	4.3	1257	US-09-173-891-6	Sequence 6, Appl
11	107	4.3	1257	US-09-076-137-6	Sequence 6, Appl
12	107	4.3	1257	PCT-US92-03624-6	Sequence 6, Appl
13	104.5	4.2	737	US-08-119-361-5	Sequence 5, Appl
14	104.5	4.2	737	US-08-336-308A-4	Sequence 4, Appl
15	104.5	4.2	737	US-08-822-324-4	Sequence 4, Appl
16	104.5	4.2	737	US-09-450-931-4	Sequence 4, Appl
17	104.5	4.2	1687	US-08-570-311-29	Sequence 29, Appl
18	104.5	4.2	1704	US-08-336-308A-10	Sequence 10, Appl
19	104.5	4.2	1704	US-08-822-324-6	Sequence 6, Appl
20	104.5	4.2	1704	US-09-490-931-10	Sequence 10, Appl
21	103.5	4.1	1290	US-08-470-350B-2	Sequence 2, Appl
22	102	4.1	433	US-09-104-623A-4	Sequence 4, Appl
23	102	4.1	635	US-08-873-479-43	Sequence 43, Appl
24	101.5	4.1	746	PCT-US95-10509-2	Sequence 2, Appl
25	97.5	3.9	713	US-08-346-237-9	Sequence 9, Appl
26	96.5	3.9	481	US-08-943-714-10	Sequence 10, Appl
27	96	3.8	532	US-08-579-777A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-426-072-2
; Sequence 2, Application US/09426072

; Patent No. 6146869

; GENERAL INFORMATION:

; APPLICANT: Paul Harris

; APPLICANT: Kimberly M. Brown

; TITLE OF INVENTION: Polypeptides Having Phospholipase B

; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same

; FILE REFERENCE: 5951.000-US

; CURRENT APPLICATION NUMBER: US/09/426,072

; CURRENT FILING DATE: 1999-10-21

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 464

; TYPE: PRT

; ORGANISM: Aspergillus oryzae

US-09-426-072-2

Query Match 100.0%; Score 2501; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 9.7e-235;
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKSTALLTGLGLASGLASPVTSYTSVREAPFGYKPGSKESIKKDKVENIWLILE	60
Db	1	MKSTALLTGLGLASGLASPVTSYTSVREAPFGYKPGSKESIKKDKVENIWLILE	60
Qy	61	NRSFNILLGVRROGLDNPINNGPFCNKNASDPSSGKYCTQAKDYDSVNDPDHVSVTGN	120
Db	61	NRSFNILLGVRROGLDNPINNGPFCNKNASDPSSGKYCTQAKDYDSVNDPDHVSVTGN	120
Qy	121	NLEFGYTPNNGAIAAGKVVADQSGFLNAQNDYPKLAPEATQVMGYTTEEEVPTLV	180
Db	121	NLEFGYTPNNGAIAAGKVVADQSGFLNAQNDYPKLAPEATQVMGYTTEEEVPTLV	180
Qy	181	DLVDEFTTFNFWSCVPGPTNPNRLCALAGTAAGHKNDLFLNYGSKSIFEAANEKG	240
Db	181	DLVDEFTTFNFWSCVPGPTNPNRLCALAGTAAGHKNDLFLNYGSKSIFEAANEKG	240
Qy	241	VSMNLNDGTNGEFPDLSLFTTYNQTSSNNVVFENFQDAYLGVLPKFSINPSCCGTN	300
Db	241	VSMNLNDGTNGEFPDLSLFTTYNQTSSNNVVFENFQDAYLGVLPKFSINPSCCGTN	300
Qy	301	TNSMHPGTNVSVEFVKQIYDAIRQGPQWDKTLFLFTYDGTGGFVDHVPPLAVRPDLN	360
Db	301	TNSMHPGTNVSVEFVKQIYDAIRQGPQWDKTLFLFTYDGTGGFVDHVPPLAVRPDLN	360
Qy	361	TYTETAKNGQKTYLHFDRLGRRMPTWVISPYSKKGYIEQIGTDPVTGKPAFYSATSVLKT	420

Db 605 IKDLYYSEKASGPAMTYVPVFAAAG-LLNHGSSQSYLSKSVLPYLRAAPAFSEQ 663
Qy 218 NDDDFLNGISSKI-FAANEK-----GVSWLNYDGTNGEPEDSLFTYVNTSRS 269
Db 664 SDDNFTNGLTOPAPPLTANGFLOSLFGLTGIRY---SYEVDPDF---KKINRLRFP 717
Qy 270 NVYPVENFODAYLGVLP-----KFSYINPSCGNTNSMHPNTGVSVEFVKQIYDA 323
Db 718 N--PIE-----LPLPGGIALRNFKMNP-----739
Qy 324 IROGQWOKTLLFTYDETGGFYDH-----VPPPLAVRPNLTYTETAKNGOKYT 373
Db 740 -----VLDIIDDHNGTIVHKSGDVPPIHKIPNRSLIHQDQINFNGSSENERKN 789
Qy 374 LH---FDRGLGRMPTWISPSKKGVIQYGT 402
Db 790 LERDDVRGDPM-----RMDRYGT 809
RESULT 4
; Patent No. 5223424
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
; RICHARD D.
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
; ACID SEQUENCE
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/225,032
; FILING DATE: 27-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,519
; FILING DATE: 27-JUL-1987
; APPLICATION NUMBER: 933,107
; FILING DATE: 20-NOV-1986
; APPLICATION NUMBER: 902,887
; FILING DATE: 02-SEP-1986
; APPLICATION NUMBER: 887,140
; FILING DATE: 17-JUL-1986
; APPLICATION NUMBER: 823,102
; FILING DATE: 27-JAN-1986
; APPLICATION NUMBER: 773,430
; FILING DATE: 06-SEP-1985
; SEQ ID NO:13:
; LENGTH: 579
5223424-13

Query Match 4.4%; Score 109.5; DB 6; Length 579;
Best Local Similarity 21.4%; Pred. No. 0.051;
Matches 86; Conservative 44; Mismatches 151; Indels 121; Gaps 22;

Qy 110 FNDDSHVTCNNL-----EFY-----GTYP-----NNGAIAKGVADQS 145
Db 121 FNAWOLISNMTTEINLVSEQIEFNVVYLKTFIESATSPSKIYNNDLTSLMVALDTN 180
Qy 146 GFLMAQNDYPKLAPEEATROVMGYTTEEVPTLVDLDEFTTNSWFSVCPGPNRNL 205
Db 181 NTL-----PYTPAARSET---LGFY---PWLPT-----KPTQRYLLSCIRNLNPP--- 221
Qy 206 CALAGTAAGHKNDODFLNYGSISSKIFEAANEKGVSWLVYDG---TNG--EFEPDSLFF 260
Db 222 -----YTGSGQIITDSIQGLHSDIMFTTIENTIENAVP IHLTGTDEFSTGIYHFDKPKL 276
Qy 261 TYVNTSRNVVVENFFQ-----DAYLGVLPKFSYINPSCGNT-NSMHPNTGVSVE 314
Db 277 THSNQTRSLGLPPKVLPEPTTEGQDHPGTLP-----GANTRKGYHOTINNSY 326
Qy 315 VFVKQIYDAIROGP-QWOKTLLFTYDETGGFYDHVPPPLAVRPNLTYTETAKNGOKYT 373
Db 327 A-----TALRPAQGVNTPYMNFDSYNGGPF-----TPIVPTADTYQYDDPEPNAIRFT 376

Qy 374 L-----HFDRGLGRMPTWISPSKKGVI---EQYG-----TDPVTGKP 409
Db 377 MGYQHGLTTSQELERYTFNPSKCGRAPKQFNOQAPNLNNTNNGTLLPSPDPIGGKS 436
Qy 410 -----APYSA---TSVLKTLGYLWDIE---DFTPRV 434
Db 437 NKHEMTLNTYGPLTALNNTAPVFPNGQIWDKELDTDLAPRL 478
RESULT 5
US-08-049-783-2
; Sequence 2, Application US/08049783
; Patent No. 5439881
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E
; APPLICANT: Schwab, George E
; APPLICANT: Payne, Jewel M
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049.783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2 (PS33F2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS33F2
; IMMEDIATE SOURCE:
; CLONE: PS33F2a
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1257
US-08-049-783-2

Query Match 4.3%; Score 107; DB 1; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

Qy 6 LLTGGLGLASLG-LASPVTSYTSVREAPFGKPGSKESLENKDKVENIWLLEN--- 61
Db 74 VLSVLTILPEVGTVAASASTIVSFIPKIFGDKPNKAKNIFEELKPOLEALIQODITNQD 133
Qy 62 ----RSEFNLILGVRRQGLDNPINNGFCNKNASDPSSGKYCTQARDYDSVFN----- 112
Db 134 AINOKFDSLQKTLNLYTV---AIDNDYVTAK-----TQLENLSILTSISIF 180

QY 113 -PDHVSVTGNLLEFYGTTPNNGAIAAGKVV-ADQSGFLNAQLNDYPKLAPEEATQVMGY 170
Db 181 IPEGYETG-GLPYAVANAHILLRDAIVNAEKLGSDEKVDTHKKY-----IKMTIHN 234
QY 171 YTEEEVPTLVDLVEFTTF--NSWFSVCPGPTNRLCALAGTAAGHCKNDODPLANGIS 228
Db 235 HTEAVTKAFGLDOKFKSLDVNSY-----YINPSCCGT-----NKNANY--- 264
QY 229 SKSIFEAANEKGVSNLYDGTNGEPEPSLFFTYVN-----QTSRSNVPVVENF 277
Db 265 IKGTEMWLDLVALWPTFDPDHYQKEVELEFTRTISPIYQVPKMNQNTSSIVPSDLF 324
QY 278 FQDAYLGVLPK--FS-----YINPSCCGT-----NKNANY--- 264
Db 325 H---YOGDLVKLEFSTRDNDGLAKIFGIRNTFYKSPNTHETIYHVDYSYNTQS---SGN 378
QY 310 VSYG-----EVEVKQIYD-----AIRQGPQMDKTLF 336
Db 379 ISRGSSNPIDLNNPIIISTCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP----- 430
QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKTYLHFDRLGGRMPTWVI 388
Db 431 -----TGGAWDHSFESDGAPEGHKLNYYTSPGDLRDFINVTLLISTPTINELSTEKI 485
QY 389 SPY-SKKGYIEQYGTDPVTGKP 409
Db 486 KGFPAAEKGYIKNOGIMKYGKP 507

RESULT 6

US-08-158-232-6
; Sequence 6, Application US/08158232
; Patent No. 5596071

GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Ulick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,232
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCU104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1257 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: PS33F2
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC2316) B-18785
FEATURE:
NAME/KEY: Protein
CLONE: 1..1257
LOCATION: 1..1257
US-08-158-232-6

Query Match 4.3%; Score 107; DB 1; Length 1257;

Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LITGLGLLASLG-LASPVTSYTSVREAPFGYKPGSKESIENLKDKVENIVWLILEN--- 61
Db 74 VLSVLTLPYEGTVASAASSTIVSFWPKIFGDKPNKAKNIFEELKQIEALIQDITNYQD 133
QY 62 ----RSFNILGGVRRROGLDNPINNGPCNKNASDPSSSKYCTQAKDYDSVFN----- 112
Db 134 AINQKFKDSLOKTLNLYTV--AIDNNYVTAK-----TOLENNLSILTSISIF 180
QY 113 -PDHVSVTGNLLEFYGTTPNNGAIAAGKVV-ADQSGFLNAQLNDYPKLAPEEATQVMGY 170
Db 181 IPEGYETG-GLPYAVANAHILLRDAIVNAEKLGSDEKVDTHKKY-----IKMTIHN 234
QY 171 YTEEEVPTLVDLVEFTTF--NSWFSVCPGPTNRLCALAGTAAGHCKNDODPLANGIS 228
Db 235 HTEAVTKAFGLDOKFKSLDVNSY-----NKNANY--- 264
QY 229 SKSIFEAANEKGVSNLYDGTNGEPEPSLFFTYVN-----QTSRSNVPVVENF 277
Db 265 IKGTEMWLDLVALWPTFDPDHYQKEVELEFTRTISPIYQVPKMNQNTSSIVPSDLF 324
QY 278 FQDAYLGVLPK--FS-----YINPSCCGT-----NKNANY--- 264
Db 325 H---YOGDLVKLEFSTRDNDGLAKIFGIRNTFYKSPNTHETIYHVDYSYNTQS---SGN 378
QY 310 VSYG-----EVEVKQIYD-----AIRQGPQMDKTLF 336
Db 379 ISRGSSNPIDLNNPIIISTCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP----- 430
QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKTYLHFDRLGGRMPTWVI 388
Db 431 -----TGGAWDHSFESDGAPEGHKLNYYTSPGDLRDFINVTLLISTPTINELSTEKI 485
QY 389 SPY-SKKGYIEQYGTDPVTGKP 409
Db 486 KGFPAAEKGYIKNOGIMKYGKP 507

RESULT 7

US-08-304-626-6
; Sequence 6, Application US/08304626
; Patent No. 5616495
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.

Query Match 4.3%; Score 107; DB 2; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23

RESULT 9
US-08-611-928-6
; Sequence 6, Application US/08611928
; Patent No. 5824792

```
QY 6 LRLGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESIENLKDKVENIWMILLEN--- 61
Db 74 VLSVLTLPVETGTVASAASITVSIWPKIFGDKPNKAKNIFEELKQIEALIQOQDITNYQD 133
QY 62 ----RSFONILGGVRRQGLDNPINNGPCNKYKNASDPSSCKYCTQAKDYDSVND---- 112
Db 134 AINQKFKDSLOKTLNLTIV--AIDNNDVYTK-----TQLENLSILTSISIF 180
QY 113 -PDHVSVTGNLEFYGTTPNNGAIAAGKVV-ADQSGFLNAQLNDYKPLAPEATQVMGY 170
Db 181 IPEGYETG-GLPYAMVANAHILLRLDAIVNAEKLGSDEKVDTHKKY-----IKMTIHN 234
QY 171 YTEEVPTLVLDVDEFTTF--NSWFSCVPGPTNPNRLCALAGTAAAGHKNDDDLNYGIS 228
Db 235 HTEAVIKAFNLGDKFKSLDVSNI-----NKKANY--- 264
QY 229 SKSIFPAANEKGVSNLNYDGTNGEPEPSLFFTYVN-----QTSRSNVVPEVF 277
Db 265 IKGTEMVLDLVALWPTFDPDHYQKEVEIEFTRTISSPIYQVPVKNMONTSSIVPSDLF 324
QY 278 FODAYLGVLPK--FS-----YINPSCCGT-----NTNSMHPTGN 309
Db 325 H----YOGDLVKLEFSTRTDNDGLAKIFTGIRNFTYKSPNTHETHYHVSFYNTQS---SGN 378
QY 310 VSYG-----EVEFKQIYD-----AIRQGPQWDKTLF 336
Db 379 ISRGSSNPIDLANNPITSTCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP----- 430
QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLTHFDRLGGRMPTWVI 388
Db 431 ----TGGANDHSFIESDGAPEGHKLNIIYISPGDTLRDFINVTYLISTPTINELSTEKI 485
QY 389 SPY-SKKGYIEQYGTDPVTGKP 409
Db 486 KGPPAEKGYIKNOGIMYYGKP 507
.
RESULT 10
US-09-173-891-6
; Sequence 6, Application US/09173891
; Patent No. 6077937
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Kennedy, M. Keith
; APPLICANT: Randall, John Brooks
; APPLICANT: Meier, Henry
; APPLICANT: Uick, Heidi Jane
; APPLICANT: Fonceerrada, Luis
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Schwab, George E.
; APPLICANT: Fu, Jenny
; TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active
; AGAINST HYMENOPTERAN PESTS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/173,891
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/158,232
```

```
; FILING DATE:
; APPLICATION NUMBER: US 07/887,980
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/797,645
; FILING DATE: 25-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/703,977
; FILING DATE: 22-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/SCJ104.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS33P2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC2316) B-18785
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1257
; US-09-173-891-6
Query Match 4.3%; Score 107; DB 3; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;
QY 6 LRLGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESIENLKDKVENIWMILLEN--- 61
Db 74 VLSVLTLPVETGTVASAASITVSIWPKIFGDKPNKAKNIFEELKQIEALIQOQDITNYQD 133
QY 62 ----RSFONILGGVRRQGLDNPINNGPCNKYKNASDPSSCKYCTQAKDYDSVND---- 112
Db 134 AINQKFKDSLOKTLNLTIV--AIDNNDVYTK-----TQLENLSILTSISIF 180
QY 113 -PDHVSVTGNLEFYGTTPNNGAIAAGKVV-ADQSGFLNAQLNDYKPLAPEATQVMGY 170
Db 181 IPEGYETG-GLPYAMVANAHILLRLDAIVNAEKLGSDEKVDTHKKY-----IKMTIHN 234
QY 171 YTEEVPTLVLDVDEFTTF--NSWFSCVPGPTNPNRLCALAGTAAAGHKNDDDLNYGIS 228
Db 235 HTEAVIKAFNLGDKFKSLDVSNI-----NKKANY--- 264
QY 229 SKSIFPAANEKGVSNLNYDGTNGEPEPSLFFTYVN-----QTSRSNVVPEVF 277
Db 265 IKGTEMVLDLVALWPTFDPDHYQKEVEIEFTRTISSPIYQVPVKNMONTSSIVPSDLF 324
QY 278 FODAYLGVLPK--FS-----YINPSCCGT-----NTNSMHPTGN 309
Db 325 H----YOGDLVKLEFSTRTDNDGLAKIFTGIRNFTYKSPNTHETHYHVSFYNTQS---SGN 378
QY 310 VSYG-----EVEFKQIYD-----AIRQGPQWDKTLF 336
Db 379 ISRGSSNPIDLANNPITSTCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP----- 430
QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLTHFDRLGGRMPTWVI 388
Db 431 ----TGGANDHSFIESDGAPEGHKLNIIYISPGDTLRDFINVTYLISTPTINELSTEKI 485
QY 389 SPY-SKKGYIEQYGTDPVTGKP 409
```

```

; Sequence 6, Application PC/TUS9203624
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schnepf, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Focerrada, Luis
; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes
; TITLE OF INVENTION: Which Code Therefor
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/03624
; FILING DATE: 19920501
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA20C2C1C1
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6 (PS33F2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS33F2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC 2316) B-18785
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1257
; PCT-US92-03624-6

Query Match 4.3%; Score 107; DB 5; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LLTGLGLASLG-LASPTSEYTSVREAPFGYKPGSKESIEENLKDKVENIVWLILEN--- 61
DB 74 VLSVLTLEPEVGTASASTVSFIWPKIFGDKPNAKNIIEELKPKQIEALIQDITNQD 133
QY 62 ----RSPDNILGVRROGLDNPINNGPFCNYKNASDFSSGKYCTQAKDYDSVFN 112
DB 134 AINQKFKDSLQKTINLYTV--AIDNDYVTAK-----TQLENLSILTSISIF 180
QY 113 -PDHSVTGNLLEFYCTYPNNGAIGASKVY-ADQSGFLNAQLNDYKPLAPEATRVQMGY 170
DB 181 IPEGYETG-GLPYAYAMANAHIILLRDAIVNAEKIGFSDKEVDTHKKY-----IKMTIH 234
QY 171 YTEEEVPTLVLDVDEFTTF--NSWFCVPGPTNPNRLCALAGTAAGHGKNDDEFNLNGIS 228
DB 235 HTEAVIKAFINGLDKFKSLDVSNSY-----NKRANY----- 264

RESULT 11
US-09-076-137-6
; Sequence 6, Application US/09076137B
; Patent No. 6166195
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schnepf, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Focerrada, Luis
; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes Which Code
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MA-20CCCD2
; CURRENT APPLICATION NUMBER: US/09/076.137B
; CURRENT FILING DATE: 1998-05-12
; EARLIER APPLICATION NUMBER: 08/316,301
; EARLIER FILING DATE: 1994-09-30
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1257
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
; US-09-076-137-6

Query Match 4.3%; Score 107; DB 4; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LLTGLGLASLG-LASPTSEYTSVREAPFGYKPGSKESIEENLKDKVENIVWLILEN--- 61
DB 74 VLSVLTLEPEVGTASASTVSFIWPKIFGDKPNAKNIIEELKPKQIEALIQDITNQD 133
QY 62 ----RSPDNILGVRROGLDNPINNGPFCNYKNASDFSSGKYCTQAKDYDSVFN 112
DB 134 AINQKFKDSLQKTINLYTV--AIDNDYVTAK-----TQLENLSILTSISIF 180
QY 113 -PDHSVTGNLLEFYCTYPNNGAIGASKVY-ADQSGFLNAQLNDYKPLAPEATRVQMGY 170
DB 181 IPEGYETG-GLPYAYAMANAHIILLRDAIVNAEKIGFSDKEVDTHKKY-----IKMTIH 234
QY 171 YTEEEVPTLVLDVDEFTTF--NSWFCVPGPTNPNRLCALAGTAAGHGKNDDEFNLNGIS 228
DB 235 HTEAVIKAFINGLDKFKSLDVSNSY-----NKRANY----- 264
QY 229 SKSIFEAANEKGVSNLYNDGTNGEFDFSLFFTYVN-----QTSRNVVYPVENF 277
DB 265 IKGTEVMDLVALWPTDPDHYGKEVEIEFTRTISSPIQPVKPKNQMTSSSIVPSDLF 324
QY 278 FODAYLGVLPK-FS-----YINPSCCGI-----NTNSMHPGTN 309
DB 325 H----YQGDVVKLEFSTRDNDGLAKIFTGIRNFTYKSPNTHYHVDYSYFQS---SGN 378
QY 310 WSYG-----EVFKQIYD-----AIRQGPQWDKTLF 336
DB 379 ISRGSSNPIDILNPIIISIRNSFYKAIAGSSVLVNFKDGTVGFAFAQAP----- 430
QY 337 ITYDETGGFYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLHEDRLGGRPTWVI 388
DB 431 -----TGGAWDHSFIESGAGEGPKLUNYIYTPGDTLRFNIVYTLTPTNELSTETKI 485
QY 389 SPY-SKKGYIEQVDPVTKP 409
DB 486 KGFPAAEKGYIKNGIMKYGKP 507

RESULT 12
PCT-US92-03624-6
```

```

QY 229 SKSIFAANKEGVSALNYDGTNGEPEPDSLEFTYVN-----QTSRNVVPPVNEF 277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 IKGMEWLDLVALWPTDFPDHYQKEVEIEFTRISSPIVQPVPKNMQNTSSSIVPSDLF 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 FQDAYLGLVLPK--ES-----YINPSCCGT-----NTNSMPTGN 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 325 H---YOGDLVLEFSTRDMDGLAKIFTGIRNTEYKSPNTHYHVSFNTQS---SGN 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 VSYG-----EVFVKQIYD-----AIROGPQWDKTLFF 336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 ISRGSSNPIDLNLPIIISIRNFSYKAIAGSSVLNFKDGTQGYAFAPQAP----- 430
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 337 ITYDETFGGEYDH--VPPPLAVRPNLTETAKNG-----QKYLTHFDRLGGRPWTWVI 388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 -----TGGAWDHSIESDGAPEGHKLNIYTPSGDTLRDFINYYTLISTPIINELSTEKI 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 389 SPY-SKKGYIEQYGTDPVTGKP 409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 KGPPAEKGYIKNQIMKYGKP 507
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-08-119-361-5
; Sequence 5, Application US/08119361
; Patent No. 5523390
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis Arginine-specific Proteinase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,361
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 21-93
; TELEPHONE: 303-499-8080
; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-119-361-5

Query Match 4.2%; Score 104.5; DB 1; Length 737;
Best Local Similarity 20.0%; Pred. No. 0.23;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

QY 36 YKPGSKESLENLKKRVENIVWLLENRSFDNLGGVRRQGLDNPINNGPFCNYKNASDPS 95
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 YEGRYTPVE---EKQGRMIVIAKYE---GDIK-----DFYDWMKQRG-- 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 96 SGKYCTQAKDYDSVFNPDHSHVTGNLFEY--GTYTPNNGAIAAGKGVADQSGFLNAQLN 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 263 ---LRTEVK---VAEDIASPTANAIQOFVKOEYKEGNDLTYLLVGDH-----K 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 154 DYP-KLAPEEATROVMG-----YYTE-----EVPPLVDLVDDE---TFENS 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 DPAKITGCIKSDQVYQIIVGNDHYNEVFIGFSCSKEDLKTQIDRTIHYERNITTEDEK 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 WFSQVPGPTNPNRLC---ALAGTAAGCHKNDDEFLNY-----GISS 229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 WLQ-----QALCIASAEYGGPSADNGESDIQHENVIANLLTQYGYTKIIKCYDPCGVT 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 KSIFEAANEKGVSWLNDGTNGEPEPDSLFF--TYVNOTSRSNVVPVENFFQDA----- 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 KNIIDAFN--GGISLVNTYTHGSETAWGTSHFGTTHVKQLTNSQLP---FIFDVACVNGD 475
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 YLGLVLPK-----SYNPS-----CCGTWNSMHP 306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 FLFSMPCFAEALMRAQDKGPTGTVAITASTINQSNASPMRGODEMNEITCEKHPNNIKR 535
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 T-GNVSYGGEV---KQIYDAIROGPW---DKTLLFIYDETFGFDHYVPPPLAY--RP 357
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 TFGGVTHMGMEAMVEKYKDKGKMLDTWTVFGDSPSLVRL-----VPTKMQYTAP 586
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 358 DNLTYTET-----AKNGOKY-----TLHFDRLGGRMP-TWVISPY 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 587 AQINLTDSVNVSCDYNGAIAITISANGKMGFSVAVENGATATINLTGLTNESLTLYVGY 646
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 SKKGXIEQYGTDPVTGKPAPY-----SATSVLTKLGYLWD 426
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 647 NKETVIKINTIN---GEPNYPQVSNLTATTTQGGKQVTLKWD 684
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-08-336-308A-4
; Sequence 4, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089

```

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 737 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-336-308A-4

Query Match      4.2%; Score 104.5; DB 3; Length 737;
Best Local Similarity 20.0%; Pred. No. 0.23;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

QY 36 YKPGKSESTENLKDVENIWLILENRSFDNLGVRQGLDNPINNGPFCNYKNASDPS 95
DB 223 YEPGRTPVE---EKQGRMIVIVAKYE---GDIK-----DFVDWKNQRG-- 262
QY 96 SGKYCTQAKDYSDVFNPDHSHVGNLNLEFY--GYTPNNGAIASGVKVDQSGFLNAQLN 153
DB 263 ---LRTEVK---VAEDIASPVTAIAIQOVFKQYEKEGNDLTYVLLVGDH-- 307
QY 154 DYP-KLAPEEATROVMG-----YYTE-----EEVPTLVLDVDEF---TTFNS 191
DB 308 DIPAKITPGIKSDQYGVGNVDHYNVEVIFGRFSCSKEDLKTQIDRTIHYERNITTEK 367
QY 192 WFCVPGPNPNRLC---ALAGTAAGHGKNDODFLNY-----GISS 229
DB 368 WLG-----QALCIASAGGFSADNGESDQHENVIANLLTQYGTIILKCYDPGVTP 419
QY 230 KSIFEAAKEGVSWLNVDGTNGEPEPDSLFF--TYVNTSRSNVVPVENFTODA----- 281
DB 420 KNIIDAFN--GGISLVNVTGHCSETAMGTSHFCTTHVKQLTNSQLP---FIFDVACVNGD 475
QY 282 YLGVLPRF-----SYNPS-----CCGTNTNSMHP 306
DB 476 FLFSMPCFAELMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNRIK 535
QY 307 T-GNVSIGEYFV---KQIYDAIRQGPW---DKTLFLITYDETGGFYDHPVPPPLAY-RP 357
DB 536 TFGGVTMNGMFAMVEKKGKGMKLDTWVFGDPSLLVRL-----VPTKMQVTAP 586
QY 358 DNLTYTET-----AKNGQY-----TLHFDRLGGRMP-TWVISPY 391
DB 587 AQINLTDASVNSVDYNGAIATISANGKMFSGSAVVENGATINLTGLTNESTLTIVVGY 646
QY 392 SKKGYEQYGTDPVTKGPAPY-----SATSVLKTLGYLMD 426
DB 647 NKEVTKTINTN---GEPNPYQPVSNLTATTGQKQVTLKWD 684

RESULT 15
US-08-822-324-4
; Sequence 4, Application US/08822324
; Patent No. 6129917
; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822.324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 488-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 737 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-822-324-4

Query Match      4.2%; Score 104.5; DB 3; Length 737;
Best Local Similarity 20.0%; Pred. No. 0.23;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

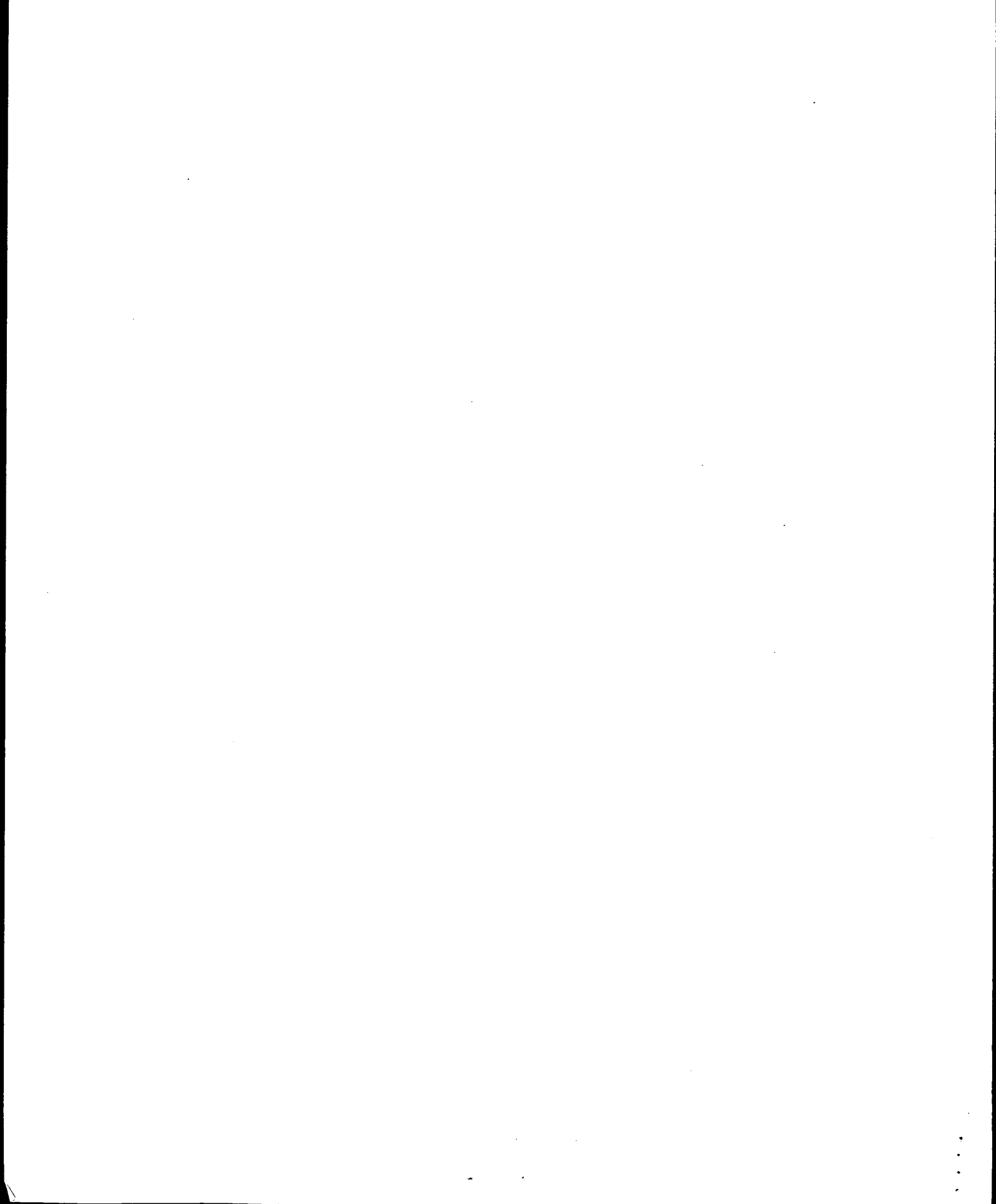
QY 36 YKPGKSESTENLKDVENIWLILENRSFDNLGVRQGLDNPINNGPFCNYKNASDPS 95
DB 223 YEPGRTPVE---EKQGRMIVIVAKYE---GDIK-----DFVDWKNQRG-- 262
QY 96 SGKYCTQAKDYSDVFNPDHSHVGNLNLEFY--GYTPNNGAIASGVKVDQSGFLNAQLN 153
DB 263 ---LRTEVK---VAEDIASPVTAIAIQOVFKQYEKEGNDLTYVLLVGDH-- 307
QY 154 DYP-KLAPEEATROVMG-----YYTE-----EEVPTLVLDVDEF---TTFNS 191
DB 308 DIPAKITPGIKSDQYGVGNVDHYNVEVIFGRFSCSKEDLKTQIDRTIHYERNITTEK 367
QY 192 WFCVPGPNPNRLC---ALAGTAAGHGKNDODFLNY-----GISS 229
DB 368 WLG-----QALCIASAGGFSADNGESDQHENVIANLLTQYGTIILKCYDPGVTP 419
QY 230 KSIFEAAKEGVSWLNVDGTNGEPEPDSLFF--TYVNTSRSNVVPVENFTODA----- 281
DB 420 KNIIDAFN--GGISLVNVTGHCSETAMGTSHFCTTHVKQLTNSQLP---FIFDVACVNGD 475
QY 282 YLGVLPRF-----SYNPS-----CCGTNTNSMHP 306
DB 476 FLFSMPCFAELMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNRIK 535
QY 307 T-GNVSIGEYFV---KQIYDAIRQGPW---DKTLFLITYDETGGFYDHPVPPPLAY-RP 357
DB 536 TFGGVTMNGMFAMVEKKGKGMKLDTWVFGDPSLLVRL-----VPTKMQVTAP 586
QY 358 DNLTYTET-----AKNGQY-----TLHFDRLGGRMP-TWVISPY 391
DB 587 AQINLTDASVNSVDYNGAIATISANGKMFSGSAVVENGATINLTGLTNESTLTIVVGY 646
QY 392 SKKGYEQYGTDPVTKGPAPY-----SATSVLKTLGYLMD 426
DB 647 NKEVTKTINTN---GEPNPYQPVSNLTATTGQKQVTLKWD 684

Search completed: December 4, 2001, 15:08:54
Job time: 114 sec
```


Wed Dec 5 09:18:17 2001

us-09-426-072-2.ra1

Page 11



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 16:58:37 ; Search time 88.78 Seconds
(without alignments)
5913.224 Million cell updates/sec

Title: US-09-426-072-1
Perfect score: 2318
Sequence: 1 aattctggacaacaatcccc.....gcaccccggtacataatt 2318

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2318	100.0	2318	3 US-09-426-072-1	Sequence 1, Appli
2	21	0.9	30	3 US-09-426-072-4	Sequence 4, Appli
3	19	0.8	35060	3 US-08-814-095-7	Sequence 7, Appli
4	18	0.8	27	3 US-09-426-072-3	Sequence 3, Appli
5	17	0.7	556	3 US-09-058-489-25	Sequence 25, Appli
6	17	0.7	1303	1 US-08-265-086-1	Sequence 1, Appli
7	17	0.7	2456	4 US-09-064-693A-19	Sequence 19, Appli
8	17	0.7	6641	4 US-09-064-693A-25	Sequence 25, Appli
9	17	0.7	49272	1 US-08-614-770A-1	Sequence 1, Appli
10	17	0.7	50341	1 US-08-247-901C-1	Sequence 1, Appli
11	17	0.7	50341	2 US-09-075-904-1	Sequence 1, Appli
12	17	0.7	52297	4 US-09-426-436-1	Sequence 1, Appli
13	17	0.7	52297	4 US-08-705-557-1	Sequence 1, Appli
14	17	0.7	80161	3 US-09-036-987A-1	Sequence 1, Appli
15	17	0.7	80161	4 US-09-370-700-1	Sequence 1, Appli
16	17	0.7	152331	3 US-09-128-155-16	Sequence 16, Appli
17	17	0.7	176373	3 US-09-128-155-17	Sequence 17, Appli
18	16	0.7	30	4 US-09-033-333-17	Sequence 17, Appli
19	16	0.7	30	4 US-09-033-428-18	Sequence 18, Appli
20	16	0.7	159	2 US-08-890-880-19	Sequence 19, Appli
21	16	0.7	159	3 US-08-890-890-19	Sequence 19, Appli
22	16	0.7	159	3 US-09-032-894-19	Sequence 19, Appli
23	16	0.7	159	4 US-09-031-626-19	Sequence 19, Appli
24	16	0.7	389	5 PCT-US95-15428-15	Sequence 15, Appli
25	16	0.7	423	1 US-08-470-179-83	Sequence 83, Appli
26	16	0.7	435	2 US-09-108-051-1	Sequence 1, Appli
27	16	0.7	435	4 US-09-440-833-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-426-072-1
; Sequence 1, Application US/09426072

; Patent No. 6146869

; GENERAL INFORMATION:

; APPLICANT: Paul Harris

; APPLICANT: Kimberly M. Brown

; TITLE OF INVENTION: Polypeptides Having Phospholipase B

; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same

; FILE REFERENCE: 5951.000-US

; CURRENT APPLICATION NUMBER: US/09/426,072

; CURRENT FILING DATE: 1999-10-21

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 2318

; TYPE: DNA

; ORGANISM: Aspergillus oryzae

US-09-426-072-1

Query Match 100.0%; Score 2318; DB 3; Length 2318;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	aattctggacaacaatcccccttttgagtgatgataaggggagtgactgagctaaag	60
Db	1	aattctggacaacaatcccccttttgagtgatgataaggggagtgactgagctaaag	60
Qy	61	ctcagctttctgcttggcgtaggctgccccatgtcttggtttagtcttctccacc	120
Db	61	ctcagctttctgcttggcgtaggctgccccatgtcttggtttagtcttctccacc	120
Qy	121	ctagtagtgactgcttgccttcccaatcccatcccatcccatcccatcccatcccat	180
Db	121	ctagtagtgactgcttgccttcccaatcccatcccatcccatcccatcccatcccat	180
Qy	181	gaccaaggctgtcaaatgagagcccccatctatgtggctcccatctatgtttttttt	240
Db	181	gaccaaggctgtcaaatgagagcccccatctatgtggctcccatctatgtttttttt	240
Qy	241	gtcgaagaagctgtcaatcccggtggtggttcccgaggaagcattgttaagaagtg	300
Db	241	gtcgaagaagctgtcaatcccggtggtggttcccgaggaagcattgttaagaagtg	300
Qy	301	atagcgatagctgcttcccttggctggtatgaagccgaagtcctctgctcgatgatat	360
Db	301	atagcgatagctgcttcccttggctggtatgaagccgaagtcctctgctcgatgatat	360
Qy	361	gcgcttcaaccgcgcacagcgcacgtggtatccgaatgggacggaaaggtagctctc	420

28	16	0.7	479	2	US-08-890-980-6	Sequence 6, Appli
29	16	0.7	479	3	US-08-890-979-6	Sequence 6, Appli
30	16	0.7	479	3	US-09-032-894-6	Sequence 6, Appli
31	16	0.7	479	4	US-09-031-626-6	Sequence 6, Appli
32	16	0.7	480	5	PCT-US96-04648-1	Sequence 1, Appli
33	16	0.7	742	4	US-09-085-848-2	Sequence 2, Appli
34	16	0.7	872	4	US-09-020-956-23	Sequence 23, Appli
35	16	0.7	872	4	US-09-030-607-23	Sequence 23, Appli
36	16	0.7	1031	1	US-08-181-271A-106	Sequence 106, App
37	16	0.7	1031	1	US-08-449-315-106	Sequence 106, App
38	16	0.7	1031	1	US-08-444-803-106	Sequence 106, App
39	16	0.7	1031	1	US-08-449-043-106	Sequence 106, App
40	16	0.7	1031	1	US-08-456-265A-106	Sequence 106, App
41	16	0.7	1031	1	US-08-455-416-106	Sequence 106, App
42	16	0.7	1031	1	US-08-455-244-106	Sequence 106, App
43	16	0.7	1031	1	US-08-454-876-106	Sequence 106, App
44	16	0.7	1031	2	US-08-457-364-106	Sequence 106, App
45	16	0.7	1031	4	US-09-350-600-106	Sequence 106, App

```
|||||
Db 361 gcgttcaaccgacgcacagcgacgctgggtatccgaatgggacgaaagtgcgtcttc 420
QY 421 cagtggtctataagccttccctggtccggtgcagactccctgacgaagcagattggt 480
Db 421 cagtggtctataagccttccctggtccggtgcagactccctgacgaagcagattggt 480
QY 481 caagctcaagcactcaatccgcgagtaaacatgaagtccaccgctctcttactggctc 540
Db 481 caagctcaagcactcaatccgcgagtaaacatgaagtccaccgctctcttactggctc 540
QY 541 ggcctcttgctctcgtctctgagcctctcagctcgagatacagaggtcgca 600
Db 541 ggcctcttgctctcgtctctgagcctctcagctcgagatacagaggtcgca 600
QY 601 gaagcccttctcgatacaagcctggtcccaaggagtgcaattgagaactggaagacaag 660
Db 601 gaagcccttctcgatacaagcctggtcccaaggagtgcaattgagaactggaagacaag 660
QY 661 gtcgagacattgtctggttatctcgagaaacaggttggtgcccctacgatactcat 720
Db 661 gtcgagacattgtctggttatctcgagaaacaggttggtgcccctacgatactcat 720
QY 721 atgtgaaataaattcctcaattcagcttggtgtgtgtgaaagcagactcaacatat 780
Db 721 atgtgaaataaattcctcaattcagcttggtgtgtgtgaaagcagactcaacatat 780
QY 781 ggtccagatcccttgataaactcttgagggtggtgcgcgcgaagactggacaaccga 840
Db 781 ggtccagatcccttgataaactcttgagggtggtgcgcgcgaagactggacaaccga 840
QY 841 tcacaagcgcgcgtctgcaactacaagaatgcgagcaccatctcgtgggaagtact 900
Db 841 tcacaagcgcgcgtctgcaactacaagaatgcgagcaccatctcgtgggaagtact 900
QY 901 gtactcaggcccaagactatgattccgtgttcaacgactcagaccactccgtgactgta 960
Db 901 gtactcaggcccaagactatgattccgtgttcaacgactcagaccactccgtgactgta 960
QY 961 ataacttgagttctacgaacttaaccccaacaatggtgcgattgcgaatggcaagg 1020
Db 961 ataacttgagttctacgaacttaaccccaacaatggtgcgattgcgaatggcaagg 1020
QY 1021 tcgtgcgcgaccagctcgtgcttccaaacgacagcttaacgactaccaccaactggccc 1080
Db 1021 tcgtgcgcgaccagctcgtgcttccaaacgacagcttaacgactaccaccaactggccc 1080
QY 1081 cagaagagcccaagcaagtgtatgagatactatcagagagaggttctcactcgtcg 1140
Db 1081 cagaagagcccaagcaagtgtatgagatactatcagagagaggttctcactcgtcg 1140
QY 1141 tggaccttggtatgagttcaacttctcaacagctggttctcgtgttctctgggta 1200
Db 1141 tggaccttggtatgagttcaacttctcaacagctggttctcgtgttctctgggta 1200
QY 1201 gtcgataaactgtatatttttaatttgactaacgcttcaaaagctaccacccca 1260
Db 1201 gtcgataaactgtatatttttaatttgactaacgcttcaaaagctaccacccca 1260
QY 1261 accgcttgctctgcttgccggaacgcgtggtggcatgggacgaatgacgatcttc 1320
Db 1261 accgcttgctctgcttgccggaacgcgtggtggcatgggacgaatgacgatcttc 1320
QY 1321 tgaactatggtatctctagcaagctccatctcgagcgcgcgaagagggcgtgctct 1380
Db 1321 tgaactatggtatctctagcaagctccatctcgagcgcgcgaagagggcgtgctct 1380
QY 1381 ggcctcaactacgattggcaccacggagaattcgaacccggtattctctcttccactag 1440
Db 1381 ggcctcaactacgattggcaccacggagaattcgaacccggtattctcttcttccactag 1440
QY 1441 tcacaagacgtcccggttcaacgctggtgcgcgttgaaacttttcaacagcgtacc 1500
|||||
```

```
Db 1441 tcacaagacgtcccggtccaaagctgtgcccgttgaaacttttccaagagcgtacc 1500
QY 1501 tcgggtctcctccctaaattctcttaccattaaacccctcctgctgcgcgaaccaacac 1560
Db 1501 tcgggtctcctccctaaattctcttaccattaaacccctcctgctgcgcgaaccaacac 1560
QY 1561 ccatgcaacccacacgcgttaacgctctcactacggtgaggtcttcgtaagcagatctatg 1620
Db 1561 ccatgcaacccacacgcgttaacgctctcactacggtgaggtcttcgtaagcagatctatg 1620
QY 1621 ccattgcccagggccctcagtgaggagaacacctgctcttcttaccctacgagcagacg 1680
Db 1621 ccattgcccagggccctcagtgaggagaacacctgctcttcttaccctacgagcagacg 1680
QY 1681 gtggtttctacgacattgtctctccctctcgcgttcgcgcgcgcgcgaacactgacctaca 1740
Db 1681 gtggtttctacgacattgtctctccctctcgcgttcgcgcgcgcgcgaacactgacctaca 1740
QY 1741 ctgagactgcgaagaacgcgttcagaaatacactcttcaactcgacgctggtggcgca 1800
Db 1741 ctgagactgcgaagaacgcgttcagaaatacactcttcaactcgacgctggtggcgca 1800
QY 1801 tgcgacgtggtttatctcccttaccagtaagaagggatacatcgagcagtaagcagacg 1860
Db 1801 tgcgacgtggtttatctcccttaccagtaagaagggatacatcgagcagtaagcagacg 1860
QY 1861 atccgctcacgggcgaagcgcgtctcactacagtgctacgtccgtccctcaagactctcgat 1920
Db 1861 atccgctcacgggcgaagcgcgtctcactacagtgctacgtccgtccctcaagactctcgat 1920
QY 1921 atctctgggacatcgagacttcaacccctcgtgtcgcccaactctcatcttcttgatcac 1980
Db 1921 atctctgggacatcgagacttcaacccctcgtgtcgcccaactctcatcttcttgatcac 1980
QY 1981 tgcgtgcgcacgacttgcgtgaggtgctctcttctcctcaagactcccatcactttt 2040
Db 1981 tgcgtgcgcacgacttgcgtgaggtgctctcttctcctcaagactcccatcactttt 2040
QY 2041 cgtgataagctcagtcgctgcagcagagattaaagtgatgaagtgatgaactctgac 2100
Db 2041 cgtgataagctcagtcgctgcagcagagattaaagtgatgaagtgatgaactctgac 2100
QY 2101 ttcagtgaacattacatgcttagagcattgttttctttagcttagctacgctagacgag 2160
Db 2101 ttcagtgaacattacatgcttagagcattgttttctttagcttagctacgctagacgag 2160
QY 2161 cgcgctgataagataaagctgggttattctctctattgtattcttatttaataagaagac 2220
Db 2161 cgcgctgataagataaagctgggttattctctctattgtattcttatttaataagaagac 2220
QY 2221 ttgtataacatgggatttaaaaggaatgtttcttgcaaatcaactcaactcagcgaacag 2280
Db 2221 ttgtataacatgggatttaaaaggaatgtttcttgcaaatcaactcagcgaacag 2280
QY 2281 cgttaactttagctgactgcaccccggtacattaatt 2318
Db 2281 cgttaactttagctgactgcaccccggtacattaatt 2318
```

RESULT 2

```
US-09-426-072-4/c
; Sequence 4, Application US/09426072
; Patent No. 6146869
; GENERAL INFORMATION:
; APPLICANT: Paul Harris
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Polypeptides Having Phospholipase B
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5951.000-US
; CURRENT APPLICATION NUMBER: US/09/426.072
; CURRENT FILING DATE: 1999-10-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 4
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-426-072-4

Query Match      0.9%; Score 21; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2030 ccataccttttcggtataagt 2050
Db 30 CCATACCTTTTCGGTATAAGT 10

RESULT 3
US-08-814-095-7/c
; Sequence 7, Application US/08814095
; Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6025183thwestern Highway, Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,095
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Montgomery, Ilene N.
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Cosmid including ACHE
; DESCRIPTION: promotor, ACHE gene and ARS gene"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 7q22
; FEATURE:
; NAME/KEY: promotor
; LOCATION: 4089..22464
; OTHER INFORMATION: /function= "ACHE Promotor"
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33297..33408)
; OTHER INFORMATION: /gene= "AR"
; FEATURE:
; NAME/KEY: 22465..22537

; OTHER INFORMATION: /function= "non-translated"
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 24090..25177
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "(translation start:
; OTHER INFORMATION: 24110)"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25524..26009
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27005..27274
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27255..28007
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 27385..27387
; FEATURE:
; NAME/KEY: exon
; LOCATION: 28008..28129
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 28129..28131
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (34528..34895)
; OTHER INFORMATION: /function= "arsenite resistance
; OTHER INFORMATION: gene"
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (34092..34358)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33779..33963)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33493..33591)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33297..33408)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 5
```

```
FEATURE:
NAME/KEY: exon
LOCATION: complement (32959..33094)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: exon
LOCATION: complement (32569..32628)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 7
FEATURE:
NAME/KEY: exon
LOCATION: complement (32386..32468)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 8
FEATURE:
NAME/KEY: exon
LOCATION: complement (31894..32080)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 9
FEATURE:
NAME/KEY: exon
LOCATION: complement (31363..31534)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 10
FEATURE:
NAME/KEY: exon
LOCATION: complement (31131..31284)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 11
FEATURE:
NAME/KEY: exon
LOCATION: complement (30816..31011)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 12
FEATURE:
NAME/KEY: exon
LOCATION: complement (30470..30626)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 13
FEATURE:
NAME/KEY: exon
LOCATION: complement (30187..30274)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 14
FEATURE:
NAME/KEY: exon
LOCATION: complement (29945..30073)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 15
FEATURE:
NAME/KEY: exon
LOCATION: complement (29664..29856)
OTHER INFORMATION: /gene="ARS"
OTHER INFORMATION: /number= 16
US-08-814-095-7
```

```
Query Match 0.8%; Score 19; DB 3; Length 35060;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 536 gtctggcctcttgccctc 554
|||||
DB 6964 GTCTGGGCTCTTGGCCTC 6946
```

```
RESULT 4
US-09-426-072-3
; Sequence 3, Application US/09426072
; Patent No. 6146869
; GENERAL INFORMATION:
; APPLICANT: Paul Harris
```

```
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Polypeptides Having Phospholipase B
; FILE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5951.000-US
; CURRENT APPLICATION NUMBER: US/09/426.072
; CURRENT FILING DATE: 1999-10-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-426-072-3
```

```
Query Match 0.8%; Score 18; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 511 atgaagtcaccgcgtctg 528
|||||
DB 10 atgaagtcaccgcgtctg 27
```

```
RESULT 5
US-09-058-489-25
; Sequence 25, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WHI97-08pa
; CURRENT APPLICATION NUMBER: US/09/058.489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041.877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-25
```

```
Query Match 0.7%; Score 17; DB 3; Length 556;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 808 gaggcgtgcgcgcgcacaa 824
|||||
DB 214 gaggcgtgcgcgcgcacaa 230
```

```
RESULT 6
US-08-265-086-1
; Sequence 1, Application US/08265086
; Patent No. 5576191
; GENERAL INFORMATION:
; APPLICANT: Gayle, Margit
; APPLICANT: Slack, Jennifer
; APPLICANT: Gross, Hans-Juergen
; APPLICANT: Sims, John E.
; APPLICANT: Dower, Steven K.
; TITLE OF INVENTION: NO. 5576191el Cytokine That Binds ST2
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
```

STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,086
FILING DATE: June 17, 1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2825
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: huST2-11g
FEATURE:
NAME/KEY: CDS
LOCATION: 88..771
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 88..156
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 157..768
US-08-265-086-1

Query Match 0.78; Score 17; DB 1; Length 1303;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 caaggagtcattgaga 646
|||||
DB 549 CAAGGAGTCATTGAGA 565

RESULT 7
US-09-064-693A-19
; Sequence 19, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/265,086
; FILING DATE: June 17, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2825
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822

COMPUTER: Toshiba Satellite Pro T2150CDS
OPERATING SYSTEM: Windows95
SOFTWARE: Word Perfect 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,693A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: W. Gary Goodson
REGISTRATION NUMBER: 22,387
REFERENCE/DOCKET NUMBER: LIT-PI-296
TELECOMMUNICATION INFORMATION:
TELEPHONE: (208)526-9469
TELEFAX: (208)526-8339
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2456 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-064-693A-19

Query Match 0.78; Score 17; DB 4; Length 2456;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2078 aaagtgtatgaacgt 2094
|||||
DB 1511 AAAGTGTATGAACGT 1527

RESULT 8
US-09-064-693A-25
; Sequence 25, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339

```

; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-25

Query Match      0.7%; Score 17; DB 4; Length 6641;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2078 aaagtatgatgaacgt 2094
DB 5696 AAAGTATGATGAACGT 5712

RESULT 9
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PROCESSOR (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,901C
; FILING DATE: MAY 23, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/057,531
; FILING DATE: APRIL 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BOGOSIAN, ELIZABETH A
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/273
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 OR 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50341
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: L5 shuttle plasmid sequence
; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: L5 mycobacteriophage
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION: No. 5750384e
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; PAGES:
; DATE:

```

```

Query Match      0.7%; Score 17; DB 1; Length 49272;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
DB 20902 TCTGGGACATCGAGGAC 20918

```


DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1

Query Match 0.7%; Score 17; DB 1; Length 50341;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
|||||
DB 20287 TCTGGGACATCGAGGAC 20303

RESULT 11

US-09-075-904-1
Sequence 1, Application US/09075904
Patent No. 5994137

GENERAL INFORMATION:
APPLICANT: Jacobs, et al.
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,904
FILING DATE: May 11, 1998

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,901
FILING DATE: May 23, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Elizabeth A
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/475
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 50341
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: L5 shuttle phasmid sequence
HYPOTHETICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: L5 mycobacteriophage
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:

POSITION IN GENOME:
CHROMOSOME/SEGMENT:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5994137e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1

Query Match 0.7%; Score 17; DB 2; Length 50341;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
|||||
DB 20287 TCTGGGACATCGAGGAC 20303

RESULT 12

US-09-426-436-1
Sequence 1, Application US/09426436
Patent No. 6225066
GENERAL INFORMATION:
APPLICANT: William R. Jacobs, Jr.
APPLICANT: Barry R. Bloom
APPLICANT: Graham F. Hatfull
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/426,436
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/705,557
FILING DATE:
APPLICATION NUMBER: US/08/057,531
FILING DATE:
APPLICATION NUMBER: 07/833,431
FILING DATE: February 7, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: 96700/238
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52297
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: phage genome sequence
; DESCRIPTION: phage genome sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: not applicable.
; ORIGINAL SOURCE:
; ORGANISM: mycobacteriophage L5
; STRAIN: not applicable
; DEVELOPMENTAL STAGE: not applicable
; HAPLOTYPE: not applicable
; TISSUE TYPE: not applicable
; CELL LINE: not applicable
; ORGANELLES: not applicable
; IMMEDIATE SOURCE: mycobacteriophage L5 particles
; POSITION IN GENOME: entire genome
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Hatfull and Sarkis
; TITLE: DNA Sequence, Structure and Gene
; TITLE: Expression of Mycobacteriophage L5:
; TITLE: A Phage System for Mycobacterial
; TITLE: Genetics
; JOURNAL: Molecular Microbiology
; VOLUME: 7
; PAGES: 395-405
; DATE: 1993
; US-09-426-436-1

Query Match 0.7%; Score 17; DB 4; Length 52297;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1924 tctgggacatcgaggac 1940
Db 20184 TCTGGGACATCGAGGAC 20200

RESULT 13
US-08-705-557-1
; Sequence 1, Application US/08705557
; Patent No. 6300061
; GENERAL INFORMATION:
; APPLICANT: William R. Jacobs, Jr.
; APPLICANT: Barry R. Bloom
; APPLICANT: Graham F. Hatfull
; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52297
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: phage genome sequence
; DESCRIPTION: phage genome sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: not applicable.
; ORIGINAL SOURCE:
; ORGANISM: mycobacteriophage L5
; STRAIN: not applicable
; DEVELOPMENTAL STAGE: not applicable
; HAPLOTYPE: not applicable
; TISSUE TYPE: not applicable
; CELL LINE: not applicable
; ORGANELLES: not applicable
; IMMEDIATE SOURCE: mycobacteriophage L5 particles
; POSITION IN GENOME: entire genome
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Hatfull and Sarkis
; TITLE: DNA Sequence, Structure and Gene
; TITLE: Expression of Mycobacteriophage L5:
; TITLE: A Phage System for Mycobacterial
; TITLE: Genetics
; JOURNAL: Molecular Microbiology
; VOLUME: 7
; PAGES: 395-405
; DATE: 1993
; US-09-426-436-1

APPLICATION NUMBER: US/08/705,557
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/057,531
FILING DATE:
APPLICATION NUMBER: 07/833,431
FILING DATE: February 7, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: 96700/238
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 52297
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: phage genome sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: not applicable.
ORIGINAL SOURCE:
ORGANISM: mycobacteriophage L5
STRAIN: not applicable
INDIVIDUAL ISOLATE: L5
DEVELOPMENTAL STAGE: not applicable
HAPLOTYPE: not applicable
TISSUE TYPE: not applicable
CELL LINE: not applicable
ORGANELLES: not applicable
IMMEDIATE SOURCE: mycobacteriophage L5 particles
POSITION IN GENOME: entire genome
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Hatfull and Sarkis
TITLE: DNA Sequence, Structure and Gene
TITLE: Expression of Mycobacteriophage L5:
TITLE: A Phage System for Mycobacterial
TITLE: Genetics
JOURNAL: Molecular Microbiology
VOLUME: 7
PAGES: 395-405
DATE: 1993
US-09-426-436-1

Query Match 0.7%; Score 17; DB 4; Length 52297;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1924 tctgggacatcgaggac 1940
Db 20184 TCTGGGACATCGAGGAC 20200

RESULT 14
US-09-036-987A-1/c
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
```

APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 0.7%; Score 17; DB 3; Length 80161;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1013 tggcaaggtcgtcgccg 1029
|||||

Db 48698 TGGCAAGGTCGTGCGCG 48682

RESULT 15

US-09-370-700-1/c
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J
APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 80161

; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 0.7%; Score 17; DB 4; Length 80161;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1013 tggcaaggtcgtcgccg 1029
|||||

Db 48698 TGGCAAGGTCGTGCGCG 48682

Search completed: December 4, 2001, 18:38:47
Job time: 6010 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 15:08:31 ; Search time 1960.54 Seconds
(without alignments)
12705.028 Million cell updates/sec

Title: US-09-426-072-1

Perfect score: 2318
Sequence: 1 aattctggacaacaatccc.....gcaccccgatacattaat 2318

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	23	1.0	472	10	BE586656 WHE0509_A
2	21	0.9	533	13	AQ706057 HS_5550_A
3	20	0.9	420	13	AZ701728 RPCI-23-1
4	20	0.9	452	13	AQ995660 RPCI-23-3
5	20	0.9	464	11	BI322812 kx16e05.Y
6	20	0.9	530	10	AI323058 m51h08.Y
7	19	0.8	138	10	AA953538 on80d06.S
8	19	0.8	151	10	AW081841 xb55h09.X
9	19	0.8	159	10	AW664184 h105f07.X
10	19	0.8	164	11	BI063061 IL3-UT011
11	19	0.8	178	13	AZ491723 1M0325024
12	19	0.8	205	11	BF543903 UI-R-Y0-a

13	19	0.8	212	10	AW176306
c 14	19	0.8	220	10	AI203319
c 15	19	0.8	220	10	AI695918
c 16	19	0.8	223	11	BI061646
17	19	0.8	223	11	BI063217
18	19	0.8	223	11	BI063298
19	19	0.8	223	11	BI063327
c 20	19	0.8	223	11	BI063392
21	19	0.8	227	11	N56396
22	19	0.8	229	11	BI062159
c 23	19	0.8	229	11	BI062404
24	19	0.8	229	11	BI062466
25	19	0.8	229	11	BI063909
c 26	19	0.8	237	11	BI061709
c 27	19	0.8	238	10	AI197858
c 28	19	0.8	239	11	BI063783
c 29	19	0.8	247	11	BI061955
c 30	19	0.8	247	11	BI062259
31	19	0.8	247	11	BI062982
c 32	19	0.8	248	11	BI037052
c 33	19	0.8	255	11	BI063770
34	19	0.8	260	10	BB350572
35	19	0.8	266	13	AZ908773
c 36	19	0.8	272	10	AA652961
c 37	19	0.8	280	11	BI010685
c 38	19	0.8	282	10	AA927182
c 39	19	0.8	291	10	AA635619
c 40	19	0.8	295	11	T35546
41	19	0.8	296	10	BB347933
42	19	0.8	299	10	AI909043
c 43	19	0.8	302	10	AA809401
c 44	19	0.8	304	11	D69313
c 45	19	0.8	322	11	225166

ALIGNMENTS

RESULT 1

BE586656

LOCUS

DEFINITION WHE0509_A09_A172R Secale cereale aluminum-stressed root tip cDNA library Secale cereale cDNA clone WHE0509_A09_A17, mRNA sequence.

ACCESSION BE586656

VERSION BE586656.1

KEYWORDS GI:9839688

SOURCE EST.

ORGANISM rye.

REFERENCE Secale cereale

AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Secale.

TITLE 1 (bases 1 to 472)

JOURNAL Anderson,O.D., Butler,E., Chao,S., Gustafson,J.P., Han,P.S., Hsiao,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.

COMMENT The structure and function of the expressed portion of the wheat genomes - Aluminum-stressed root tip cDNA library from rye (Secale cereale)

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: M13 reversed primer.

Location/Qualifiers

1..472

/organism="Secale cereale"

/cultivar="Blanco"

```

/db_xref="taxon:4550"
/clone="WHE0509_A09_A17"
/clone_lib="Secale cereale aluminum-stressed root tip cDNA
library"
/tissue_type="Root tip"
/dev_stage="Seedling"
/lab_host="E. coli DH12S"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Seeds
were germinated aseptically on filter paper and
transferred to a hydroponic growth system in a growth
chamber when the primary root was 1 cm in length. After a
2-day establishment period, seedlings were subjected to a
5 ppm aluminum stress prior to tissue harvest. Plants
were grown in an environmental chamber. The tissue, total
RNA, and poly(A) RNA were prepared, and a cDNA library
was made (Butler and Gustafson) at University of Missouri
, Columbia. Plasmid DNA preparations and DNA sequencing
were performed in the OD Anderson lab (all other authors
)."
BASE COUNT      94 a 164 c 120 g  94 t
ORIGIN

```

```

Query Match      1.0%; Score 23; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.62; 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0;
QY 1680 ggtggtcttacgaccatgtccc 1702
      |||||
DB 346 GGTGGCTTCTACGACCATGTCCC 368

```

```

RESULT 2
AQ706057/c
LOCUS      AQ706057      533 bp      DNA      GSS      07-JUL-1999
DEFINITION HS_5550_A1_C07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1126 Col=13 Row=E, DNA sequence.
ACCESSION  AQ706057
VERSION    AQ706057.1 GI:5415483
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 533)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
            99380589
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3687
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1126 row: E column: 13
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 533.
            Location/Qualifiers
            1..533
            /organism="Homo sapiens"
            /db_xref="taxon:9606"

```

```

/clone="Plate=1126 Col=13 Row=E"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="Male"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT      183 a  81 c  74 g 181 t 14 others
ORIGIN
Query Match      0.9%; Score 21; DB 13; Length 533;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2222 ttgataacatgggttttaaaa 2242
      |||||
DB 99 TTGATAACATGGGATTTAAAA 79

```

RESULT 3

AZ701728/c

LOCUS AZ701728 420 bp DNA GSS 24-JAN-2001

DEFINITION RPCI-23-19611.TJ RPCI-23 Mus musculus genomic clone RPCI-23-19611,
 DNA sequence.

ACCESSION AZ701728

VERSION AZ701728.1 GI:12424417

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 420)
 Zhao,S., Nieman,W., Feldblyum,T., Malek,J., Shatsman,S., Aknret
 and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-19611.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choir.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 196 row: I column: 1
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..420
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-19611"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 115 a 95 c 105 g 105 t
 ORIGIN

Query Match 0.9%; Score 20; DB 13; Length 420;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 gatgaactgagatgctaagc 61
 |||||
 Db 189 GATGAAGTGAAGTGAAGC 170

RESULT 4
 LOCUS AQ995660 452 bp DNA 24-FEB-2000
 DEFINITION RPCI-23-355M11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-355M11
 VERSION AQ995660
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 452)
 AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akimret
 , B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-23-355M11.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 355 row: M column: 11
 Seq primer: SP6
 Class: BAC ends.

FEATURES source
 1. 452
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-355M11"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 93 a 50 c 98 g 211 t
 ORIGIN

Query Match 0.9%; Score 20; DB 13; Length 452;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 cttcatattgtttttttgt 242
 |||||
 Db 317 CTTCAVATTTGTTTCTTGT 336

RESULT 5
 LOCUS BI322812/c 464 bp mRNA 30-JUL-2001
 DEFINITION Kx16e05.y3 Parastrongyloides trichosuri IL pAMP1 vl Chiapelli
 McCarter Parastrongyloides trichosuri cDNA 5' similar to TR:O62515
 O62515 PROBABLE PEROXISOMAL PHYTANOYL-COA ALPHA-HYDROXYLASE ;, mRNA
 sequence.

ACCESSION BI322812
 VERSION BI322812.1 GI:15001998
 KEYWORDS EST.
 SOURCE Parastrongyloides trichosuri.
 ORGANISM Parastrongyloides trichosuri
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
 Panagrolaimoidea; Strongyloidea; Parastrongyloidea.

REFERENCE 1 (bases 1 to 464)
 AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
 Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
 Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., R.
 , Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
 , M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.
 , Shih, T., Jackson, I., Cardenas, M., McCann, R., Waterston, R. and
 Wilson, R.
 TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 The library was constructed by Brandi Chiapelli and Dr. James
 McCarter (bchiapell@wustl.edu & jmcarter@wustl.edu) at
 Washington University, St. Louis. DNA Sequencing by: Washington
 University Genome Sequencing Center St. Louis.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 270.

FEATURES source
 1. 464
 /organism="Parastrongyloides trichosuri"
 /db_xref="taxon:131310"
 /clone_lib="Parastrongyloides trichosuri IL pAMP1 vl
 Chiapelli McCarter"
 /dev_stage="Infective Larvae"
 /lab_host="DH10B"
 /note="Vector: pAMP1 (Gibco); Site_1: NotI; Site_2: Sall;
 The library was constructed by Brandi Chiapelli and Dr.
 James McCarter at Washington University, St. Louis. The
 cDNA was made by using Dynabead oligo-dt priming (Dynal).
 PCR based library using a modified protocol from the
 SMART PCR cDNA Synthesis Kit from Clontech. Directionally
 cloned into the UDG sites of pAMP1. Nematodes were
 provided by Dr. Warwick Grant of AgResearch, New Zealand
 (warwick.grant@agresearch.co.nz)."

BASE COUNT 185 a 62 c 69 g 148 t
 ORIGIN

Query Match 0.9%; Score 20; DB 11; Length 464;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 ttcatattgtttttttgtc 243
 |||||
 Db 68 TTCATATTGTTTCTTGTTC 49

RESULT 6
 LOCUS AI323058 530 bp mRNA 23-DEC-1998
 DEFINITION mJ51h08.y1 Soares mouse embryo NME13.5 14.5 Mus musculus cDNA

clone IMAGE:479679 5' similar to gb:X73960 M.musculus mRNA for TIE receptor tyrosine kinase (MOUSE);, mRNA sequence.

ACCESSION
AI323058
VERSION
AI323058.1 GI:4057487
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 530)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:290423

This read is a RESEQUENCE of a previously sequenced mouse clone correct orientation)

Putative full length read
vector to vector length is 613

Seq primer: -40RP from Gibco
High quality sequence stop: 325.

FEATURES
source

1. 530
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:479679"
/lab_lib="Soares mouse embryo NDME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker: Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAGTGGGAGCGCGCGGAAATTTTGTGTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 114 a 151 c 148 g 116 t 1 others
ORIGIN

Query Match 0.9%; Score 20; DB 10; Length 530;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 ctttcctcggcggtcag 457
|||||

Db 481 CTTTCCCTGGCGGCTCAG 462

RESULT 7
AA953538/c
LOCUS

AA953538 mRNA EST 07-MAY-1998

DEFINITION
ACCESSION
AA953538
VERSION
AA953538.1 GI:3117685
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 138)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 131.
Location/Qualifiers
1. 138
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1562987"
/lab_host="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"

FEATURES
source

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker: Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
25 a 41 c 27 g 45 t

Query Match 0.8%; Score 19; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 agattataaagtgtatg 2089
|||||

Db 40 AGGATTAAAGTGATGATG 22

RESULT 8
AW081841/c
LOCUS

DEFINITION
AW081841 151 bp mRNA EST 14-OCT-1999
xb55h09.x1 NCI-CGAP_Eso2 Homo sapiens cDNA clone IMAGE:2580257 3', mRNA sequence.

ACCESSION
AW081841
VERSION
AW081841.1 GI:6036993
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 151)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Nan Hu, M.D., Ph.D., Mark Roth, M.D., Phillip Taylor, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
 Seq primer: -400P from Gibco
 High quality sequence stop: 118.

FEATURES

source
 1..151
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI-CGAP_Eso2"
 /tissue_type="squamous cell carcinoma"
 /lab_host="DH10B"
 /note="Organ: esophagus; Vector: pCMV-SPORT6; Site_1: SalI
 ; Site_2: NotI; Cloned unidirectionally. Primer: Oligo
 dT. Average insert size 1.1 kb. Life Technologies catalog
 #: 11502-010"
 24 a 41 c 25 g 61 t

BASE COUNT

ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 151;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggattataaagtgtatg 2089

|||||

Db 62 AGGATTAAGTGTATG 44

RESULT 9

AW664184/c 159 bp mRNA EST 06-APR-2000
 LOCUS hi05f07.x1 NCI-CGAP_GUL Homo sapiens cDNA clone IMAGE:2971429 3',
 mRNA sequence.

ACCESSION AW664184

VERSION AW664184.1 GI:7456725

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 159)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/resources.shtml

Possible reversed clone: polyT not found

Seq primer: -400P from Gibco.

FEATURES

source
 1..159
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI-CGAP_GUL"
 /tissue_type="2 pooled high-grade transitional cell
 tumors"

/lab_host="DH10B"

/note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
 Site_1: SalI; Site_2: NotI; Cloned unidirectionally.

Primer: Oligo dT. Library constructed by Life

Technologies."

BASE COUNT 29 a 44 c 34 g 52 t
 ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 159;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggattataaagtgtatg 2089

|||||

Db 39 AGGATTAAGTGTATG 21

RESULT 10

BI063061

LOCUS IL3-UT0117-280301-538-D04 UT0117 Homo sapiens cDNA, mRNA sequence.

ACCESSION BI063061

VERSION BI063061.1 GI:14470588

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 164)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=IL3&t2=IL3-UT0117-280301-528-D04&t3=2001-03-28&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 164.

Location/Qualifiers

1..164

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0117"

/dev_stage="Adult"

/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 61 a 36 c 43 g 24 t

ORIGIN

Query Match 0.8%; Score 19; DB 11; Length 164;
 Best Local Similarity 100.0%; Pred. No. 84;

Query Match 0.8%; Score 19; DB 11; Length 205;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1169 caacagctggtctcgtgt 1187
 |||
 Db 78 CAACAGCTGGTCTCGGT 96

RESULT 13
 AW176306 212 bp mRNA EST 16-NOV-1999
 LOCUS QV0-BT0229-310899-011-d10 BT0229 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW176306
 VERSION AW176306.1 GI:6442343
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil.
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t3=QV0-BT0229-310899-011-d10&t3=1999-08-31&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 31
 High quality sequence stop: 212.

FEATURES
 source
 1. 212
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0229"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: pUC18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 62 a 64 c 48 g 38 t
 ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 212;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 614 gatacaagcctggtcccaa 632
 |||
 Db 11 GATACAAGCCTGGCTCCAA 29

RESULT 14
 AI203319/c 220 bp mRNA EST 14-OCT-1998
 LOCUS q129g06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:1942330 3',
 mRNA sequence.
 ACCESSION AI203319
 VERSION AI203319.1 GI:3755925

KEYWORDS
 SOURCE

ORGANISM

REFERENCE
 AUTHORS

TITLE

JOURNAL

COMMENT

EST.
 human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 220)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco.

Location/Qualifiers
 1. 220

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI CGAP GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
 from the normalized library NCI_CGAP_GC4 was prepared, and
 ss circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clonids
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 35 a 62 c 45 g 78 t
 ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 220;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggattaaagtgtgtgtg 2089
 |||
 Db 42 AGGATTAAAGTGATGATG 24

RESULT 15
 AI695918/c 220 bp mRNA EST 03-JUN-1999
 LOCUS ts87e09.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2238280 3',
 mRNA sequence.

ACCESSION AI695918
 VERSION AI695918.1 GI:4983818
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 220)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.

cdNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cdNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco.

FEATURES
source
1. .220
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2238280"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones IDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 35 a 62 c 45 g 78 t
ORIGIN

Query Match 0.8% Score 19; DB 10; Length 220;
Best Local Similarity 100.0%; Fred. No. 87;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2071 aggattaaaagtgtgatg 2089
|||||
Db 42 AGGATTAAAAGTGTGTG 24

Search completed: December 4, 2001, 17:35:19
Job time: 8808 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 17:03:47 ; Search time 208.51 Seconds
(without alignments)
9530.861 Million cell updates/sec

Title: US-09-426-072-1
Sequence: 1 aattctgacacacatcccc.....gcaccccgatcataatt 2318

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	2318	100.0	2318	22 AAA54535
2	21	0.9	30	22 AAA54537
3	20	0.9	3748	22 AAH54226
4	19	0.8	304	20 AAZ77482
5	19	0.8	1108	20 AAX60581
6	19	0.8	1263	21 AAH27058
7	19	0.8	1281	21 AAA27059
8	19	0.8	3778	22 AAH16638
9	19	0.8	119950	20 AAH90201
10	18	0.8	27	22 AAA54536
11	18	0.8	40	21 AAC61921

12	18	0.8	1070	22 AAH33057
c 13	18	0.8	1224	21 AAC61895
14	18	0.8	1307	21 AAZ29801
15	18	0.8	1394	22 AAH28830
c 16	18	0.8	1549	21 AAC69521
17	18	0.8	1766	21 AAH88846
18	18	0.8	1766	21 AAZ45385
19	18	0.8	1766	22 AAH27105
c 20	18	0.8	2761	22 AAH27105
21	18	0.8	4450	22 AAC61960
c 22	18	0.8	6832	22 AAH21139
c 23	18	0.8	8358	22 AAH21140
24	18	0.8	38186	20 AAH22028
25	18	0.8	38186	22 AAC90085
c 26	17	0.7	160	21 AAC18914
27	17	0.7	172	20 AAH86664
c 28	17	0.7	211	16 AAH20160
c 29	17	0.7	252	22 AAH83697
c 30	17	0.7	253	22 AAH83030
c 31	17	0.7	300	21 AAH00141
c 32	17	0.7	345	22 AAH82747
33	17	0.7	373	19 AAH28605
34	17	0.7	375	21 AAH46651
c 35	17	0.7	404	22 AAH82773
36	17	0.7	404	22 AAH67364
c 37	17	0.7	412	22 AAH83539
c 38	17	0.7	429	22 AAH66595
c 39	17	0.7	470	14 AAQ39686
c 40	17	0.7	470	14 AAQ59098
41	17	0.7	482	21 AAC75557
42	17	0.7	500	21 AAC78019
43	17	0.7	500	22 AAH33259
44	17	0.7	509	20 AAH41959
45	17	0.7	513	22 AAH67214

ALIGNMENTS

RESULT 1
AAA54535
ID AAA54535 standard; DNA; 2318 BP.
XX
AC AAA54535;
XX
DF 11-APR-2001 (first entry)
XX
DE Phospholipase B coding sequence of Aspergillus oryzae.
XX
KW Phospholipase B; hydrolysis; hydrolyse; lysophospholipid;
KW phospholipid; lecithin; lysolethacin; degumming; starch;
KW wheat starch; oil; dough; baking; strength; elasticity;
KW stability; extensibility; ds.
XX
OS Aspergillus oryzae.
XX
FH Key Location/Qualifiers
FT exon 510..696
FT FT /*tag= a
FT FT /*label= Exon 1
FT FT 697..788
FT FT /*tag= b
FT FT /*label= Intron 1
FT FT 789..1196
FT FT /*tag= c
FT FT /*label= Exon 2
FT FT 1197..1247
FT FT /*tag= d
FT FT /*label= Intron 2
FT FT 1248..2048
FT FT /*tag= e
FT FT /*label= Exon 3
XX

PS Claim 8; Page 1180-1182; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. *epidermidis* polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the

CC treatment of S. *epidermidis* infections, e.g. endocarditis. AAH53971 to

CC AAH55090 represent specifically claimed S. *epidermidis* genomic DNA

CC polynucleotide sequences from the present invention. AAH55091 to

CC AAH55098 represent oligonucleotide sequences and primers which are used

CC in the amplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,

CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 3748 BP; 1287 A; 569 C; 650 G; 1242 T; 0 other;

SQ

Query Match 0.9%; Score 20; DB 22; Length 3748;

Best Local Similarity 100.0%; Pred. No. 5.5;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 220 ccacttcattgttttctt 239

Db 2250 CCACCTCATATGTTTCTT 2231

RESULT 4

AAZ77482

ID AAZ77482 standard; cDNA; 304 BP.

XX AAZ77482;

AC AAZ77482;

DT 10-APR-2000 (first entry)

XX Human ovarian tumor cDNA library derived EST fragment 33.

DE Expressed sequence tag; EST; human; ovarian tumor; anticancer;

KW gene therapy; treatment; ss.

XX Homo sapiens.

OS

PN DE19817557-Al.

XX

PD 21-OCT-1999.

XX

PF 09-APR-1998; 98DE-1017557.

XX

PR 09-APR-1998; 98DE-1017557.

XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;

PI

PI WPI: 1999-591920/51.

DR P-PSDB; AA76578, AA76579, AA76580.

XX

XX New nucleic acid sequences expressed in ovarian, and some other, cancer

PT tissues, and derived polypeptides, for treatment of ovarian cancer and

PT identification of therapeutic agents

XX

PS Claim 3; Page 164; 310pp; German.

XX This invention describes novel nucleic acid (cDNA) sequences (A) which

CC have anticancer activity and are highly expressed in ovarian tumor

CC tissue (and some also in testis and breast cancer tissue). The products

CC of the invention can be used for gene therapy. (A) are used (i) for

CC recombinant expression of polypeptides (B) and (ii) to isolate complete

CC genes. (B) are used (i) to identify agents suitable for treatment of

CC ovarian cancer; (ii) directly for treating this form of cancer

CC (including expression from gene therapy vectors) and (iii) for generation

CC of specific antibodies. (A) are identified by assembling ESTs (expressed

CC sequence tags) from a particular tissue type before comparison of the

CC expression patterns. This allows a significantly longer fragment of the

CC gene to be revealed, so should reduce the number of failures associated

CC with the fact that ESTs from different libraries may represent different

CC parts of the same unknown gene, distorting the estimated frequency of

CC occurrence in a particular tissue. AA77450-277572 represent the human

CC ovarian tumor cDNA library derived EST fragments described in the method

CC of the invention and encode the protein fragments represented in

CC AA76505-76638.

XX Sequence 304 BP; 88 A; 66 C; 94 G; 56 T; 0 other;

SQ

Query Match 0.8%; Score 19; DB 20; Length 304;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 807 ggaggcgtgcgcgcgaag 825

Db 207 ggaggcgtgcgcgcgaag 225

RESULT 5

AAZ60581

ID AAZ60581 standard; DNA; 1108 BP.

XX AAZ60581;

AC AAZ60581;

XX

DT 27-JUL-1999 (first entry)

XX Human secreted protein encoding DNA (clone bh157_7).

DE

XX Secreted protein; human; tissue marker; genetic disease; gene therapy;

KW veterinary medicine; cell proliferation; immunostimulant; infection;

KW immunosuppressant; autoimmune disease; organ rejection; tumour; anemia;

KW haematopoiesis; wound healing; fertility control; chemotaxis; analgesic;

XX thrombolytic; haemophilia; infarction; antimicrobial agent; cancer; ss.

XX Homo sapiens.

OS

PN WO9924469-A1.

XX

PD 20-MAY-1999.

XX

PF 06-NOV-1998; 98WO-US23829.

XX

PR 04-NOV-1998; 98US-0185936.

XX

PR 07-NOV-1997; 97US-0965789.

XX

PA (GEM) GENETICS INST INC.

XX

PI Agostino MJ, Evans C, Jacobs K, Lavallie ER, McCoy JM;

PI Merberg D, Racie LA, Treacy M;

XX

XX WPI: 1999-327362/27.

DR P-PSDB; AA76781.

XX

XX Nucleic acid encoding secreted human proteins

PT

XX Claim 19; Page 93; 107pp; English.

XX

CC The invention provides polynucleotides (AAH60579-X60687) encoding

CC specific secreted human proteins (AA76779-Y16787). The nucleic acid

CC sequences are deposited under the accession number ATCC 98580. The

CC polynucleotides are used as tissue markers, chromosomal tags, for

CC diagnosis of genetic diseases, to generate anti-protein or anti-DNA

CC antibodies, also as nutritional sources and supplements and in gene
 CC therapy. The secreted proteins are useful therapeutically, in human or
 CC veterinary medicine, e.g. for modulating cell proliferation or
 CC differentiation, as immunostimulants or immunosuppressants (for treating
 CC infections, autoimmune disease, organ rejection, or to induce tumour
 CC immunity), as regulators of haematopoiesis (e.g. for treating anemia or
 CC in conjunction with tumour therapy), to stimulate growth of tissue for
 CC wound healing, as fertility control agents, for regulating chemotaxis or
 CC chemokines (e.g. for directing cells to tumours or sites of infection), as
 CC haemostatic and thrombolytic agents (e.g. in treatment of haemophilia or
 CC infarctions), as antimicrobial agents, for modifying biorhythms,
 CC appetite, or metabolism, as analgesics and many other uses. The proteins
 CC are also used to raise antibodies, used as diagnostic immunoassay
 CC reagents also (when neutralizing) for treating e.g. cancer.
 CC
 SQ Sequence 1108 BP; 441 A; 191 C; 299 G; 177 T; 0 other;

Query Match 0.8%; Score 19; DB 20; Length 1108;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggattaaaagtgtatg 2089
 |||||
 Db 907 aggtataaagtgtatg 925

RESULT 6
 AAA27058
 ID AAA27058 standard; DNA; 1263 BP.
 XX AC AAA27058;
 XX DT 22-AUG-2000 (first entry)
 XX DE Human 5T4 tumour-associated antigen gene.
 XX KW Human; TAA; tumour-associated antigen; anti-tumour; cytostatic;
 KW immunostimulant; vaccine; carcinoma; colorectal cancer;
 KW gastric cancer; ds.
 XX OS Homo sapiens.
 XX PN WO200029428-A2.
 XX PD 25-MAY-2000.
 XX PF 18-NOV-1999; 99WO-GB03859.
 XX PR 18-NOV-1998; 98GB-0025303.
 XX PR 27-JAN-1999; 99GB-0001739.
 XX PR 30-JUL-1999; 99GB-0017995.
 XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX PI Carroll MW, Myers KA;
 XX DR WPI; 2000-387735/33.
 XX PT Tumor associated antigen, 5T4 capable of eliciting cytotoxic
 PT T-lymphocyte response useful in vaccinating against and in treating
 PT tumors -
 XX PS Example 2; Page 78; 79pp; English.

XX The present sequence encodes the human 5T4 tumour-associated
 CC antigen (TAA). The TAA 5T4 is a glycoprotein which is widely
 CC expressed in carcinomas but has a highly restricted expression pattern
 CC in normal adult tissues. It appears to be strongly correlated to
 CC metastasis in colorectal and gastric cancer. 5T4 antigen may therefore
 CC be useful in tumour diagnosis, targeting and immunotherapy.
 CC Mice in which tumours had been induced were inoculated with a virus
 CC expression vector containing the present sequence. The 5T4 antigen
 CC was shown to be effective at eliciting an immunotherapeutic anti-tumour
 CC response. Both the nucleic acid encoding the antigen and the antigen
 CC itself can be used to elicit an immune response, preferably CTL or an
 CC antibody response in a subject. The present sequence appears in
 CC GenBank at accession number AF012160.
 XX
 SQ Sequence 1281 BP; 246 A; 410 C; 354 G; 271 T; 0 other;

CC was shown to be effective at eliciting an immunotherapeutic anti-tumour
 CC response. Both the nucleic acid encoding the antigen and the antigen
 CC itself can be used to elicit an immune response, preferably CTL or an
 CC antibody response in a subject.
 XX
 SQ Sequence 1263 BP; 230 A; 428 C; 349 G; 256 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 1263;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttcttggaacaacatccct 21
 |||||
 Db 868 ttcttggaacaacatccct 886

RESULT 7
 AAA27059
 ID AAA27059 standard; DNA; 1281 BP.
 XX AC AAA27059;
 XX DT 22-AUG-2000 (first entry)
 XX DE Mouse 5T4 tumour-associated antigen gene.
 XX KW Mouse; TAA; tumour-associated antigen; anti-tumour; cytostatic;
 KW immunostimulant; vaccine; carcinoma; colorectal cancer;
 KW gastric cancer; ds.
 XX OS Mus musculus.
 XX PN WO200029428-A2.
 XX PD 25-MAY-2000.
 XX PF 18-NOV-1999; 99WO-GB03859.
 XX PR 18-NOV-1998; 98GB-0025303.
 XX PR 27-JAN-1999; 99GB-0001739.
 XX PR 30-JUL-1999; 99GB-0017995.
 XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX PI Carroll MW, Myers KA;
 XX DR WPI; 2000-387735/33.
 XX PT Tumor associated antigen, 5T4 capable of eliciting cytotoxic
 PT T-lymphocyte response useful in vaccinating against and in treating
 PT tumors -
 XX PS Example 2; Page 78; 79pp; English.

XX The present sequence encodes the mouse 5T4 tumour-associated
 CC antigen (TAA). The TAA 5T4 is a glycoprotein which is widely
 CC expressed in carcinomas but has a highly restricted expression pattern
 CC in normal adult tissues. It appears to be strongly correlated to
 CC metastasis in colorectal and gastric cancer. 5T4 antigen may therefore
 CC be useful in tumour diagnosis, targeting and immunotherapy.
 CC Mice in which tumours had been induced were inoculated with a virus
 CC expression vector containing the present sequence. The 5T4 antigen
 CC was shown to be effective at eliciting an immunotherapeutic anti-tumour
 CC response. Both the nucleic acid encoding the antigen and the antigen
 CC itself can be used to elicit an immune response, preferably CTL or an
 CC antibody response in a subject. The present sequence appears in
 CC GenBank at accession number AF012160.
 XX
 SQ Sequence 1281 BP; 246 A; 410 C; 354 G; 271 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 1281;

0:

Qy 1208 aaatctgtatatttta 1226
 |||||
 Db 3945 AAATCTGTATATTTTAA 3927

RESULT 10
 AAA54536
 ID AAA54536 standard; DNA; 27 BP.
 AC AAA54536;
 XX
 XX 11-APR-2001 (first entry)
 XX
 DE Primer for amplifying phospholipase B coding sequence.
 XX
 KW Phospholipase B; hydrolysis; hydrolyse; lysophospholipid;
 KW phospholipid; lecithin; lysolecithin; degumming; starch;
 KW wheat starch; oil; dough; baking; strength; elasticity;
 KW stability; extensibility; primer; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN US6146869-A.
 XX
 PD 14-NOV-2000.
 XX
 PF 21-OCT-1999; 99US-0426072.
 XX
 PR 21-OCT-1999; 99US-0426072.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 PI Harris P, Brown KM;
 XX
 DR WPI; 2001-049057/06.
 XX
 PT Novel polypeptide having phospholipase B activity useful for
 PT hydrolysing fatty acyl groups of phospholipid, preparing dough or baked
 PT product and reducing phospholipid content in an edible oil
 XX
 PS Example 3; Column 31; 24pp; English.
 XX
 CC Phospholipase B is useful for hydrolysing the fatty acyl group(s) of
 CC phospholipids or lysophospholipids, such as lecithin or lysolecithin,
 CC degumming an aqueous carbohydrate solution or slurry to improve its
 CC filterability, particularly, a starch hydrolysate or wheat starch
 CC hydrolysate which is difficult to filter and to reduce the
 CC phospholipid content in an edible oil by treating the oil with
 CC phospholipase B to hydrolyse a major portion of the phospholipid
 CC and separating an aqueous phase containing the hydrolysed
 CC phospholipid from the oil. Phospholipase B is also useful for
 CC preparing a dough or a baked product having improved property such
 CC as increased strength, elasticity, stability and extensibility. Two
 CC primers (AAA54536, AAA54537) were used to amplify the phospholipase B
 CC genomic DNA sequence.
 XX
 SQ Sequence 27 BP; 8 A; 6 C; 5 G; 8 T; 0 other;

Query Match 0.8%; Score 18; DB 22; Length 27;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 atgaagtcaccgcctctg 528
 |||||
 Db 10 atgaagtcaccgcctctg 27

RESULT 11
 AAC61921
 ID AAC61921 standard; DNA; 40 BP.
 XX
 AC AAC61921;

XX 06-MAR-2001 (first entry)
 DT PCR primer for DNA encoding a prenyltransferase designated ATPT3.
 XX
 DE
 XX
 KW Prenyltransferase: ATPT1; ATPT2; ATPT3; ATPT4; ATPT5; ATPT6; ATPT7;
 KW ATPT8; ATPT9; ATPT10; ATPT11; ATPT12; tocopherol; homogentisic acid;
 KW phytylpyrophosphate; 2-methyl-6phytylbenzoquinol; antioxidant;
 KW nutritional supplement; PCR primer; ss.
 XX
 OS Arabidopsis sp.
 XX
 PN WO2000063391-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 14-APR-2000; 2000WO-US10368.
 XX
 PR 15-APR-1999; 99US-0129899.
 XX
 PR 30-JUL-1999; 99US-0146461.
 XX
 PA (CALJ) CALGENE LLC.
 XX
 XX Savidge B, Lassner MW, Weiss JD, Post-Beittenmiller D;
 PI WPI; 2000-647519/62.
 XX
 DR An isolated nucleic acid sequence encoding prenyltransferase used to
 DR transform plant cells to increase the production of tocopherols -
 XX
 PS Example 2; Page 26; 114pp; English.
 XX
 CC PCR primers AAC61920-21 were used to amplify DNA encoding a
 CC prenyltransferase. The specification describes prenyltransferases
 CC designated ATPT1, ATPT2, ATPT3, ATPT4, ATPT5, ATPT6, ATPT7, ATPT8,
 CC ATPT9, ATPT10, ATPT11, and ATPT12. The biosynthesis of alpha-tocopherol
 CC in higher plants involves the condensation of homogentisic acid and
 CC phytylpyrophosphate to form 2-methyl-6phytylbenzoquinol, which can form
 CC various tocopherols. The prenyltransferase polynucleotides are useful in
 CC transforming host cells to alter the expression of prenyltransferase in
 CC these cells. The transformed cells are used in the production of
 CC tocopherols which are of use in the pharmaceutical industry as
 CC antioxidants and also in the food industry as nutritional supplements.
 XX
 SQ Sequence 40 BP; 12 A; 9 C; 7 G; 12 T; 0 other;

Query Match 0.8%; Score 18; DB 21; Length 40;
 Best Local Similarity 100.0%; Pred. No. 59; Mismatches 0; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1474 ttgaaactcttccaag 1491
 |||||
 Db 16 ttgaaactcttccaag 33

RESULT 12
 AAH33057
 ID AAH33057 standard; cDNA; 1070 BP.
 XX
 AC AAH33057;
 XX
 XX 03-SEP-2001 (first entry)
 DT Human colon cancer antigen encoding cDNA SEQ ID NO:113.
 XX
 DE Human colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX

PD 05-APR-2001.
 XX 28-SEP-2000; 2000WO-US26524.
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI WPI: 2001-235357/24.
 XX P-PSDB; AAG73626.
 DR Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX Claim 1; Page 2290-2291; 9803pp; English.
 XX AAG732943 to AAG737195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAG73196 to AAG73204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX Sequence 1070 BP; 237 A; 292 C; 293 G; 237 T; 11 other;
 SQ

Query Match 0.8%; Score 18; DB 22; Length 1070;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1644 gacaagaccctgctcttc 1661
 DB 692 gacaagaccctgctcttc 709
 |||||||
 RESULT 13
 AAC61895/c
 ID AAC61895 standard; DNA; 1224 BP.
 XX
 AC AAC61895;
 XX
 XX 06-MAR-2001 (first entry)
 DT
 DE DNA encoding an aromatic class prenyltransferase designated ATPT3.
 XX
 XX prenyltransferase: ATPT1; ATPT2; ATPT3; ATPT4; ATPT5; ATPT6; ATPT7;
 KW ATPT8; ATPT9; ATPT10; ATPT11; ATPT12; tocopherol; homogentisic acid;
 KW phytylpyrophosphate; 2-methyl-6phytylbenzoquinol; antioxidant;
 KW nutritional supplement; ss.
 XX
 OS Arabidopsis sp.
 XX
 XX Key Location/Qualifiers
 FH 1..1224
 FT CDS /*tag= a
 FT /product= "prenyltransferase ATPT3"
 FT
 XX

PN WO200063391-A2.
 XX 26-OCT-2000.
 PD 14-APR-2000; 2000WO-US10368.
 XX 15-APR-1999; 99US-0129899.
 PR 30-JUL-1999; 99US-0146461.
 XX (CALJ) CALGENE LLC.
 XX Savidge B, Lassner MW, Weiss JD, Post-Beittenmiller D;
 PI WPI: 2000-647519/62.
 XX P-PSDB; AAB19409.
 DR An isolated nucleic acid sequence encoding prenyltransferase used to
 PT transform plant cells to increase the production of tocopherols -
 XX Example 1; Page 76; 114pp; English.
 XX The present sequence encodes a prenyltransferase. The specification
 CC describes prenyltransferases designated ATPT1, ATPT2, ATPT3, ATPT4,
 CC ATPT5, ATPT6, ATPT7, ATPT8, ATPT9, ATPT10, ATPT11, and ATPT12. The
 CC biosynthesis of alpha-tocopherol in higher plants involves the
 CC condensation of homogentisic acid and phytylpyrophosphate to form
 CC 2-methyl-6phytylbenzoquinol, which can form various tocopherols. The
 CC prenyltransferase polynucleotides are useful in transforming host cells
 CC to alter the expression of prenyltransferase in these cells. The
 CC transformed cells are used in the production of tocopherols which are
 CC of use in the pharmaceutical industry as antioxidants and also in the
 CC food industry as nutritional supplements.
 XX Sequence 1224 BP; 308 A; 220 C; 288 G; 408 T; 0 other;
 SQ

Query Match 0.8%; Score 18; DB 21; Length 1224;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1474 tttgaaacctcttccaaag 1491
 DB 1221 TTGAAACCTTCTTCCAAAG 1204
 |||||||
 RESULT 14
 AAZ29801
 ID AAZ29801 standard; DNA; 1307 BP.
 XX
 AC AAZ29801;
 XX
 XX 27-MAR-2000 (first entry)
 DT
 DE Sequence downstream to Ds insert in tomato mutant UQ13.
 XX
 XX Tomato mutant UQ13; UDP-glucose-pyrophosphorylase gene; expansin gene;
 KW PMGS; phenotype modifying genetic sequence; plant pathogen resistance;
 KW senescence timing; starch metabolism; ds.
 XX
 OS Lycopersicon esculentum.
 XX
 XX WO9963068-A1.
 XX 09-DEC-1999.
 PD
 XX 04-JUN-1999; 99WO-AU00434.
 PF
 XX 04-JUN-1998; 98AU-0003901.
 PR 04-JUN-1998; 98AU-0003903.
 PR 25-SEP-1998; 98AU-0006169.
 PR 25-SEP-1998; 98AU-0006174.
 XX (UYQU) UNIV QUEENSLAND.
 PA

XX Carroll BJ;
 XX WPI; 2000-116368/10.
 XX
 XX New polynucleotides that increase gene expression in plants used to
 PT produce transgenic plants with resistance to plant pathogens -
 XX
 XX Disclosure: Page 67; 93pp; English.
 XX
 XX The present sequence is the downstream portion of Ds insert containing
 CC PMS in tomato mutant UO13. The Ds insertion is associated with a
 CC putative UDP-glucose-pyrophosphorylase and/or expansin gene, which are
 CC potentially involved in starch metabolism. This resulted in nos: BAR
 CC expression. This can be used to produce transgenic plants having altered
 CC phenotypic traits, such as resistance to plant pathogens, senescence
 CC timing, starch metabolism, cell growth, expansion and/or division, and
 CC the shape of cells, tissues or organs.
 XX
 XX Sequence 1307 BP; 497 A; 163 C; 170 G; 471 T; 6 other;
 XX
 XX
 XX Query Match 0.8%; Score 18; DB 21; Length 1307;
 XX Best Local Similarity 100.0%; Pred. No. 58;
 XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1212 ctgtatatttttaatttt 1229
 XX |||||||||||||||
 XX Db 780 ctgtatatttttaatttt 797
 XX
 XX RESULT 15
 XX AAI62830
 XX ID AAI62830 standard; DNA; 1394 BP.
 XX -XX
 XX AC AAI62830;
 XX
 XX DT 22-OCT-2001 (first entry)
 XX
 XX -DE Human genomic DNA SEQ ID NO 158.
 XX
 XX Human; neotropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
 KW ds.
 XX
 XX Homo sapiens.
 XX
 XX WO2001155449-A1.
 XX
 XX 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001WO-US01346.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 XX 04-FEB-2000; 2000US-0180628.
 XX 19-MAY-2000; 2000US-0205515.
 XX 07-JUL-2000; 2000US-0216880.
 XX 14-JUL-2000; 2000US-0218290.
 XX 14-AUG-2000; 2000US-0225447.
 XX 01-SEP-2000; 2000US-0229343.
 XX 06-SEP-2000; 2000US-0230437.
 XX 08-SEP-2000; 2000US-0231243.
 XX 25-SEP-2000; 2000US-0234997.
 XX 29-SEP-2000; 2000US-0236367.
 XX 13-OCT-2000; 2000US-0239837.
 XX 08-NOV-2000; 2000US-0246476.
 XX 08-NOV-2000; 2000US-0246477.
 XX 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246528.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249265.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-476225/51.
 XX
 XX Novel plasma membrane associated proteins useful for diagnosing,
 PT treating, preventing and/or prognosing disorders related to the
 PT proteins, including cancer, immune response and neuronal disorders -
 XX
 XX Example 2; SEQ ID NO 158; 532pp + Sequence Listing; English.
 XX
 XX The invention relates to novel genes (AAI62752-AAI62961) and proteins
 CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1394 BP; 313 A; 385 C; 338 G; 358 T; 0 other;
 XX

Query Match 0.8%; Score 18; DB 22; Length 1394;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1213 tgtatatttttaattttg 1230
 |||||||||||||||
 Db 1340 tgtatatttttaattttg 1357

Search completed: December 4, 2001, 18:37:46
 Job time: 5639 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 15:38:16 ; Search time 3277.91 Seconds
(without alignments)
11666.111 Million cell updates/sec

Title: US-09-426-072-1

Perfect score: 2318

Sequence: 1 attctcggacacaaatcccc.....gcaccccggaatacattatt 2318

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:**

1: gb.ba:**

2: gb.htg:**

3: gb.in:**

4: gb.om:**

5: gb.ov:**

6: gb.pat:**

7: gb.ph:**

8: gb.pl:**

9: gb.pr:**

10: gb.ro:**

11: gb.sts:**

12: gb.sy:**

13: gb.un:**

14: gb.vi:**

15: gb.ba:**

16: em.fun:**

17: em.hum:**

18: em.in:**

19: em.om:**

20: em.or:**

21: em.ov:**

22: em.pat:**

23: em.ph:**

24: em.pl:**

25: em.ro:**

26: em.sts:**

27: em.sy:**

28: em.un:**

29: em.vi:**

30: em.htgo.hum:**

31: em.htgo.inv:**

32: em.htgo.rod:**

33: em.htg.hum:**

34: em.htg.inv:**

35: em.htg.rod:**

36: em.htg.other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2318	100.0	2318	6	AR141633	AR141633 Sequence
2	2318	100.0	2318	6	AX127521	AX127521 Sequence
3	22	0.9	10807	1	AE004899	AE004899 Pseudomon
4	22	0.9	10807	2	AC021564	AC021564 Homo sapi
5	21	0.9	30	6	AR141633	AR141633 Sequence
6	21	0.9	30	6	AX127521	AX127521 Sequence
7	21	0.9	38627	9	AC011538	AC011538 Homo sapi
8	21	0.9	40611	9	AC021090	AC021090 Homo sapi
9	21	0.9	105538	2	AC092279	AC092279 Homo sapi
10	21	0.9	153173	2	AC091983	AC091983 Homo sapi
11	21	0.9	184606	2	AC072046	AC072046 Homo sapi
12	21	0.9	200689	2	AC067865	AC067865 Homo sapi
13	21	0.9	201312	2	AC067900	AC067900 Homo sapi
14	21	0.9	209910	2	AC073534	AC073534 Homo sapi
15	20	0.9	240	8	CNS01BDF	AL114091 Botrytis
16	20	0.9	636	8	CNS01CHX	AL115549 Botrytis
17	20	0.9	2215	3	S39048	S39048 knob associ
18	20	0.9	2521	9	BC007685	BC007685 Homo sapi
19	20	0.9	3748	1	AF269548	AF269548 Staphyloc
20	20	0.9	3748	6	AX144868	AX144868 Sequence
21	20	0.9	10645	1	AE008335	AE008335 Agrobacte
22	20	0.9	11093	1	AE005134	AE005134 Halobacte
23	20	0.9	12029	3	AE001372	AE001372 Plasmodi
24	20	0.9	63385	9	HSJ474G15	AL132673 Human DNA
25	20	0.9	80765	9	AC004009	AC004009 Homo sapi
26	20	0.9	102393	8	AC000107	AC000107 Genomic s
27	20	0.9	114364	9	AC011666	AC011666 Human Chr
28	20	0.9	126228	9	AL513211	AL513211 Human DNA
29	20	0.9	167147	2	AC026860	AC026860 Homo sapi
30	20	0.9	170594	2	AP003385	AP003385 Homo sapi
31	20	0.9	170956	2	AC024973	AC024973 Homo sapi
32	20	0.9	172805	9	AC066608	AC066608 Homo sapi
33	20	0.9	175084	2	AL161636	AL161636 Homo sapi
34	20	0.9	175233	2	AC022941	AC022941 Homo sapi
35	20	0.9	202760	2	AC021987	AC021987 Homo sapi
36	20	0.9	202924	2	AC079240	AC079240 Homo sapi
37	20	0.9	203257	9	AC013737	AC013737 Homo sapi
38	20	0.9	212494	2	AC074037	AC074037 Mus muscu
39	19	0.8	287	10	AF045954	AF045954 Mus muscu
40	19	0.8	304	6	AX014844	AX014844 Sequence
41	19	0.8	1138	9	HSA012584	AJ012584 Homo sapi
42	19	0.8	1231	9	BC000805	BC000805 Homo sapi
43	19	0.8	1260	1	AF314196	AF314196 Pseudomon
44	19	0.8	1263	6	AX025011	AX025011 Sequence
45	19	0.8	1281	6	AX025012	AX025012 Sequence

ALIGNMENTS

RESULT	1	AR141633	2318 bp	DNA	PAT	08-AUG-2001
LOCUS		Sequence 1	from patent	US 6146869.		
DEFINITION		AR141633				
ACCESSION		AR141633				
VERSION		AR141633.1	GI:15101149			
KEYWORDS		Unknown.				
SOURCE		Unknown.				
ORGANISM		Unclassified.				
REFERENCE		1 (bases 1 to 2318)				
AUTHORS		Harris, P. and Brown, K.M.				
TITLE		Polypeptides having phospholipase B activity and nucleic acids				
JOURNAL		Patent: US 6146869-A 1 14-NOV-2000;				
FEATURES		Location/Qualifiers				
source		l..2318				
BASE COUNT		554 a 655 c 534 g 575 t				
ORIGIN		/organism="unknown"				

Query Match									
Best Local Similarity 100.0%; Score 2318; DB 6; Length 2318;									
Matches 2318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	aattcctggacacaaatcccttgaagatgaatgaaggagatgaactgaagatgctaaag	60						
Db	1	AATTCTCGACAAATPCCCTTTGAGATGATAGTAAGGGGATGAACCTGAGATGCTAAG	60						
Qy	61	ctcagcttcttcgcttggctaggtgccccatgtcttgggttagtgcctctctccacc	120						
Db	61	CTCAGCTTCTCGCTTGGCTAGGCTGCCCATGTCTGGTTAGTGTCTCTCCACC	120						
Qy	121	ctagtagttagcattgctgttcaatcccatcccatcccatcccatcccatctattat	180						
Db	121	CTAGTAGTTAGCATTTGCTGTCCAATPCCCATTCCTCCCATCTCCCATCTATTAT	180						
Qy	181	gaccaaggctgcaaatgagagcccccatcatgtggcctccactcatattgtttcttt	240						
Db	181	GACCAAGGCTGTCAATGAGAGCCCCACTCATGTGGCTCCACTTCATATTGTTCTTT	240						
Qy	241	gtcgataagactgtccatccctgtccgtggttccacgaagaacattgttaaatgag	300						
Db	241	GTGATAGAGCTGTCAATPCCCTTCCGTGCTCCGTGCTTCCACGAAGAATTTGTAAAGATG	300						
Qy	301	atagcgataggctgcttggctggtgatgaagccgaagtccctgctgctcgatgat	360						
Db	301	ATAGCGATAGGCTGCCCTTGGTGGTATGAAGCCGAAGTCCCTCGCTCGATGATAT	360						
Qy	361	ggcttcaaccgacgacagcagcagcgtggtatccgaatggagcgaaggtgagcttc	420						
Db	361	GGCTTCAACCCGACGACAGCGACGTGTATCCGANTGGACGGAAAGGTAGTCTCTC	420						
Qy	421	cadtgggctataagccttccctggcgggtcagactccctgagcgaagcagctgtt	480						
Db	421	CAGTGGGTATATAAGCTTTCCCTGGCCGGGCTCAGACTCCCTAGCAAGCAAGTGT	480						
Qy	481	caagctcagcactcaatccgagatgaacatgaagtccacgcctctctactgcttc	540						
Db	481	CAGCTCAGCAGCTCAATCCGGAGTAAACATGAAGTCCACCCCTCTCTACTGTCTC	540						
Qy	541	ggccttggcctctcgttctgctagcctgtacgttcgagatatacagagctgaga	600						
Db	541	GGCTTTGGCTCTCTCGGTCTTCTAGCCCTGTACAGTCCAGCTTACAGCGGTGCGA	600						
Qy	601	gaagcccttctcgatatacaagcctgctcccaagaggtccattgagaacttgaagacaag	660						
Db	601	GAAGCCCTTTCCGATACAAAGCTGGCTCCCAAGGAGTCCATTGAGAACTTGAAGGACAAG	660						
Qy	661	gtcgaaacattgtctggcttctctcgagaaacagggttgtgccccttaagatattcat	720						
Db	661	GTGAGAACATTTCTGGCTTATTTCTCGAGAACAGSTTTGTGTGCCCTTAGCATATTAT	720						
Qy	721	atgtgaataataaatccctcaattcagcttctggtatgtgaagcagcagcactaaatat	780						
Db	721	ATGTGAATATAAATCCCTCAATTCAGCTTGTGTATGTGAAGAGGAGCACTAAATAT	780						
Qy	781	ggctcagactctcgataacattctgggagcgtgcccgcgaagaagactggaacccga	840						
Db	781	GGTCCAGATCTTTCGATAAATTTCTGGAGGCGTGGCCGCCCAAGGACTGGACACCCGA	840						
Qy	841	tcaacacggccgcttctgcaactacaagaatgagagcaccatcctcgggcaagtaact	900						
Db	841	TCAACAACGGCCGTTCTGCAACTACAAGATGGAGAGCACCATCTCTGGGCAAGTACT	900						
Qy	901	gtactcagcgaaggaactatgattccgtgttcaacgattcaagcactccgtgactggta	960						
Db	901	GTAATCAGGCCAAGGACTATGATTTCCGTGTTCACAGATCCAGACCACTCCGTGACTGGTA	960						
Qy	961	ataacttgggttctcaggaacttacaccccccaaatggtgcatgtccagtggaag	1020						
Db	961	ATAACTTGGAGTTCTACGGAACTTTACACCCCAAAATGGTGGGATTGCCAGTGGCAAG	1020						


```
QY 1501 tcggtgtctccctaaattcttcttaacattacccctctctgtcgccgacccaacaaact 1560
|||||
Db 1501 TCGGTGTCTCCCTAAATTCCTTACATTAACCCCTCTCTGCTCGCGCACCAACCAACT 1560
|||||
QY 1561 ccattgacccaccggttaacgtctctcagctgaggtctctcgaagcagatctatgatg 1620
|||||
Db 1561 CCATGACCCACCAGTAAAGTCTCTCAGGTGAGGTCTCTGTCACAGCAGATCTATGATG 1620
|||||
QY 1621 ccattgacccagggccctcagtggaagaagaccctctctctctattcattcagcagacag 1680
|||||
Db 1621 CCATTCGCCAGGGCCCTCAGTGGGACAAAGACCCCTGCTCTTCATTAACCTACGACGAGAC 1680
|||||
QY 1681 gtgctctctcagacatgtccctccctctcgcgtccgctcgcccgagacacacctgacctaca 1740
|||||
Db 1681 GTGGCTCTACGACCATGTCCCTCCCTCTCGCCGTCGCCCGGACCACTGACCTACA 1740
|||||
QY 1741 ctgagactcggaagaacggtcagaataacactcttcaacttcagccgtctgggtggcgca 1800
|||||
Db 1741 CTGAGACTCGGAAGAACGGTCAAGAAATACACTTCTACTTCGACCCGCTCGGTGGCGCA 1800
|||||
QY 1801 tgcgacctgggttatctccctctacagtaagaaggatatactcagcagcagcagacag 1860
|||||
Db 1801 TGCGACCTGGGTTATCTCCCTTACAGTAAGAAGGATACATCGAGCAGTACGGAACGG 1860
|||||
QY 1861 atccctcaggggaacgcccgtctccctacagtgtaactccgtctcctcaagactctcgat 1920
|||||
Db 1861 ATCCCGTCAAGGGCAAGCCGCTCCCTACAGTGCTACTCCGTCCTCAAGACTCTCGGAT 1920
|||||
QY 1921 atctctgggacatcgagacttcaacctctgctcgcccaactctccatctcttcgatacc 1980
|||||
Db 1921 ATCTCTGGGACATCGAGACTTCAACCTCTGCTCGGCCACTCTCCATCTTTTCGATCACC 1980
|||||
QY 1981 tgatcgacacacttgcgtgagatgctctctctctctcctcagactcccatcctttt 2040
|||||
Db 1981 TGATCGGACAGACTTGGTGAGATGCTCTCTCTCTCTCAAGACTCCCATACCTTTT 2040
|||||
QY 2041 cgggtataagctcagtcgctgagcagcagcagcagcagcagcagcagcagcagcagcag 2100
|||||
Db 2041 CGGTATAAGTCTCAGTCGCTGAGTGAAGAGGATTAAGAGTGAAGTGAAGTGAAGTGAAG 2100
|||||
QY 2101 ttcagtgacgattacattgattagagcattgttctctctctctctctctctctctctct 2160
|||||
Db 2101 TTCTGTCAGACGATTACATGTTATAGACATGTTTCTCTTATAGTACGCTAGAGCGAG 2160
|||||
QY 2161 cggatgatgataagataagcgtgggttattctctctctctctctctctctctctctctct 2220
|||||
Db 2161 CGCGATGATGATAAGATGAAGCTGGGTATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
|||||
QY 2221 ttgataacatgggtatttaaaaggaatgtttctctctctctctctctctctctctctct 2280
|||||
Db 2221 TTTGATAACATGGGATTAAAGGAATGTTTCTTGTGACATCACTCACTCACTCACTCACT 2280
|||||
QY 2281 cgtactttacatgactgaccccgagatacaatt 2318
|||||
Db 2281 CGTAACCTTACGATGCTGACCCCGGATACATTAATT 2318
|||||

RESULT 3
AE004899/c
LOCUS AE004899 10807 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 460 of 529 of the complete genome.
ACCESSION AE004899 AE004091
VERSION AE004899.1 GI:9951126
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa.
Bacteria: Proteobacteria; gamma subdivision, Pseudomonadaceae;
Pseudomonas.
1. (bases 1 to 10807)
REFERENCE
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Huftnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Salier,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
1. .10807
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
83..1372
/gene="purD"
/notes="PA4856"
83..1372
/gene="purD"
/codon_start=1
/transl_table=11
/product="phosphoribosylamine--glycine ligase"
/protein_id="AAG08240.1"
/db_xref="GI:9951127"
/translation="MNVLIIGSGREHALAMKVAQDPRAKVFVAPGNATATEAKE
NVALDVALAQADPAKKNVLTIVGPEAPLVAGVDLFRERGLDFGPTAGAAQLG
SKAFTKDFLARHRTAAYRTEVEPEAPLVAGVDLFRERGLDFGPTAGAAQLG
DEAAAVRDLGAGNAGFAGSVVIEEFLLDGEAEAFIVVDGONVLPATQSDHRRVG
DGGSPNTGGMGAYSPAPVTAEVHVDVITYTVRMAEGNVIYFGIYAGLMD
KSGAPVIFNCFRGPDEPTQINVRLESLLVLEAAKALDKVETATWDRPTGVV
LAAGPGYDYGKGEVTEGLAEAAALDQFAGTAGLKDQGVITSGRVCATAGESV
SAAQOQVYRLAEKIRWNGCFYRKIDYRAIARERGES"
1476..4304
/gene="PA4856"
1476..4304
/gene="PA4856"
/codon_start=1
/transl_table=11
/product="probable sensor/response regulator hybrid"
/protein_id="AAG08241.1"
/db_xref="GI:9951128"
/translation="MVRIRIAGLVSVFLLLLGPMSPVAVDAGVSSVPLQTTATTP
SANWRLLRDESAQLRIADVLORKFPPKAKRSFTFPASPOAVWLOVLPQAKVPS
WLVIFAPRVQVLYDLYVDGQVLRDQHTGESRPPERPILPSYLFSLPVQKPMPLY
VRMTSNHPLMAWFOIDRAGLVGLEKPAYAFGLGMLLLMYNLIRFAYSRASSL
WLAHAALACAAANGLVAFWPLGLKFNQSLTADLGAALVSLMFWACSFRTGTA
ESRLNLQGEALDILAGAITAFTQQLWFWNLVILVLSLVPLTAHWHRTGTA
PARLVAGMIVFNAGFMVFLVGTQKLDQFVGLVGFVFATLGLVSLTERQ
LIQQLNQORTSEAAHTAEIQKAEFLAKISHEIRTPMVGIVGMTLLIGTPLSAQR
DYVQTHSAGNELLTINELIDISKLESQIELDEVQFDNLALIEDCLIDIFRVAEQ
RIELISFTQPVYIGGDPTRLQVRSLLDNAPKOTEEGELLVVALDQGETPRL
RIAVODSGHPDAKERALLTAELHSDPLSASKLSHGILIIAROLVRLMGGEFGI
SGSQGTTLSLTPLDQOOLENPTADGLQAGRLVVDNETCKVLQOCCSGML
NVSASGSGKEALQRLTKAHREYFDVLLDQMPGTMGLQAAKIKEDPNLNDILL
IMLTGNSNAPSKITARNAGIKRLAKPVAGITLRLATLADLAQGVGVVTKTQPAKE
AQAPSLPDSFRLVAEDNSISTKVIKMLNLIQOPTASNGEALSAMKATQVDVLV
MDCMPVLDGFSATQELRAWEHEQRPTPVVATLHLSHKEARLRLVGMGDMHAKP
VELSQLRLIAYVWGRDRRRQGDALPS"
4760..5353
/gene="PA4857"
4760..5353
/gene="PA4857"
/codon_start=1
/transl_table=11
```

```

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE
JOURNAL

FEATURES
source
1. .10807
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
83..1372
/gene="purD"
/notes="PA4856"
83..1372
/gene="purD"
/codon_start=1
/transl_table=11
/product="phosphoribosylamine--glycine ligase"
/protein_id="AAG08240.1"
/db_xref="GI:9951127"
/translation="MNVLIIGSGREHALAMKVAQDPRAKVFVAPGNATATEAKE
NVALDVALAQADPAKKNVLTIVGPEAPLVAGVDLFRERGLDFGPTAGAAQLG
SKAFTKDFLARHRTAAYRTEVEPEAPLVAGVDLFRERGLDFGPTAGAAQLG
DEAAAVRDLGAGNAGFAGSVVIEEFLLDGEAEAFIVVDGONVLPATQSDHRRVG
DGGSPNTGGMGAYSPAPVTAEVHVDVITYTVRMAEGNVIYFGIYAGLMD
KSGAPVIFNCFRGPDEPTQINVRLESLLVLEAAKALDKVETATWDRPTGVV
LAAGPGYDYGKGEVTEGLAEAAALDQFAGTAGLKDQGVITSGRVCATAGESV
SAAQOQVYRLAEKIRWNGCFYRKIDYRAIARERGES"
1476..4304
/gene="PA4856"
1476..4304
/gene="PA4856"
/codon_start=1
/transl_table=11
/product="probable sensor/response regulator hybrid"
/protein_id="AAG08241.1"
/db_xref="GI:9951128"
/translation="MVRIRIAGLVSVFLLLLGPMSPVAVDAGVSSVPLQTTATTP
SANWRLLRDESAQLRIADVLORKFPPKAKRSFTFPASPOAVWLOVLPQAKVPS
WLVIFAPRVQVLYDLYVDGQVLRDQHTGESRPPERPILPSYLFSLPVQKPMPLY
VRMTSNHPLMAWFOIDRAGLVGLEKPAYAFGLGMLLLMYNLIRFAYSRASSL
WLAHAALACAAANGLVAFWPLGLKFNQSLTADLGAALVSLMFWACSFRTGTA
ESRLNLQGEALDILAGAITAFTQQLWFWNLVILVLSLVPLTAHWHRTGTA
PARLVAGMIVFNAGFMVFLVGTQKLDQFVGLVGFVFATLGLVSLTERQ
LIQQLNQORTSEAAHTAEIQKAEFLAKISHEIRTPMVGIVGMTLLIGTPLSAQR
DYVQTHSAGNELLTINELIDISKLESQIELDEVQFDNLALIEDCLIDIFRVAEQ
RIELISFTQPVYIGGDPTRLQVRSLLDNAPKOTEEGELLVVALDQGETPRL
RIAVODSGHPDAKERALLTAELHSDPLSASKLSHGILIIAROLVRLMGGEFGI
SGSQGTTLSLTPLDQOOLENPTADGLQAGRLVVDNETCKVLQOCCSGML
NVSASGSGKEALQRLTKAHREYFDVLLDQMPGTMGLQAAKIKEDPNLNDILL
IMLTGNSNAPSKITARNAGIKRLAKPVAGITLRLATLADLAQGVGVVTKTQPAKE
AQAPSLPDSFRLVAEDNSISTKVIKMLNLIQOPTASNGEALSAMKATQVDVLV
MDCMPVLDGFSATQELRAWEHEQRPTPVVATLHLSHKEARLRLVGMGDMHAKP
VELSQLRLIAYVWGRDRRRQGDALPS"
4760..5353
/gene="PA4857"
4760..5353
/gene="PA4857"
/codon_start=1
/transl_table=11
```

```

Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z., and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 10807)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Salier,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
1. .10807
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
83..1372
/gene="purD"
/notes="PA4856"
83..1372
/gene="purD"
/codon_start=1
/transl_table=11
/product="phosphoribosylamine--glycine ligase"
/protein_id="AAG08240.1"
/db_xref="GI:9951127"
/translation="MNVLIIGSGREHALAMKVAQDPRAKVFVAPGNATATEAKE
NVALDVALAQADPAKKNVLTIVGPEAPLVAGVDLFRERGLDFGPTAGAAQLG
SKAFTKDFLARHRTAAYRTEVEPEAPLVAGVDLFRERGLDFGPTAGAAQLG
DEAAAVRDLGAGNAGFAGSVVIEEFLLDGEAEAFIVVDGONVLPATQSDHRRVG
DGGSPNTGGMGAYSPAPVTAEVHVDVITYTVRMAEGNVIYFGIYAGLMD
KSGAPVIFNCFRGPDEPTQINVRLESLLVLEAAKALDKVETATWDRPTGVV
LAAGPGYDYGKGEVTEGLAEAAALDQFAGTAGLKDQGVITSGRVCATAGESV
SAAQOQVYRLAEKIRWNGCFYRKIDYRAIARERGES"
1476..4304
/gene="PA4856"
1476..4304
/gene="PA4856"
/codon_start=1
/transl_table=11
/product="probable sensor/response regulator hybrid"
/protein_id="AAG08241.1"
/db_xref="GI:9951128"
/translation="MVRIRIAGLVSVFLLLLGPMSPVAVDAGVSSVPLQTTATTP
SANWRLLRDESAQLRIADVLORKFPPKAKRSFTFPASPOAVWLOVLPQAKVPS
WLVIFAPRVQVLYDLYVDGQVLRDQHTGESRPPERPILPSYLFSLPVQKPMPLY
VRMTSNHPLMAWFOIDRAGLVGLEKPAYAFGLGMLLLMYNLIRFAYSRASSL
WLAHAALACAAANGLVAFWPLGLKFNQSLTADLGAALVSLMFWACSFRTGTA
ESRLNLQGEALDILAGAITAFTQQLWFWNLVILVLSLVPLTAHWHRTGTA
PARLVAGMIVFNAGFMVFLVGTQKLDQFVGLVGFVFATLGLVSLTERQ
LIQQLNQORTSEAAHTAEIQKAEFLAKISHEIRTPMVGIVGMTLLIGTPLSAQR
DYVQTHSAGNELLTINELIDISKLESQIELDEVQFDNLALIEDCLIDIFRVAEQ
RIELISFTQPVYIGGDPTRLQVRSLLDNAPKOTEEGELLVVALDQGETPRL
RIAVODSGHPDAKERALLTAELHSDPLSASKLSHGILIIAROLVRLMGGEFGI
SGSQGTTLSLTPLDQOOLENPTADGLQAGRLVVDNETCKVLQOCCSGML
NVSASGSGKEALQRLTKAHREYFDVLLDQMPGTMGLQAAKIKEDPNLNDILL
IMLTGNSNAPSKITARNAGIKRLAKPVAGITLRLATLADLAQGVGVVTKTQPAKE
AQAPSLPDSFRLVAEDNSISTKVIKMLNLIQOPTASNGEALSAMKATQVDVLV
MDCMPVLDGFSATQELRAWEHEQRPTPVVATLHLSHKEARLRLVGMGDMHAKP
VELSQLRLIAYVWGRDRRRQGDALPS"
4760..5353
/gene="PA4857"
4760..5353
/gene="PA4857"
/codon_start=1
/transl_table=11
```

```

/product="hypothetical protein"
/db_xref="GI:9951129"
/translation="MHLLFVYLKMLVLYSPFFVLSFCFISLRSFGFGRDKRMAWRVA
LAALIASVALYLFGRYIFILFGITADAFRIGAGSVLFIASLMAOGRAGVDSNVQOD
VTIPLTIPLVGPTICALLVMGCGEAKLISLVAIFLACLILGTLTILSDRIER
LLDQGLQIVSRIMGLFVCAALAAQIIVGLKSLFPG"
/length=5778
/genes="PA4858"
/length=5778
/genes="PA4858"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="GI:9951130"
/translation="MKRRSLKKAFTLSAFIASMGLSWSIOAETIKVKILHLSLGTWA
ISLTKDMALMTIDEINAKGVNGRKLEAVVDPASNPFLFAKARQLLTQDKVAVV
FGWTSVRKSVLPVFFELNGLLFPVQYEGEELSPNVFTYGAAPNQAQIPAVIYLS
EDGGARFFLLGTDIVYPRITNKILRAFLLHSGKGVDDKIDQEVYTPFGHSDYDITLAD
IKFAAGGKTAVSVTVNGDSNPVFKELANOCLEATEVPVAVFSGEELRGIDTKPL
VGLAANNVSVSNPNVEFVADKAYAKAKNLPNGTAVTNDPMEATYVGIHWAQ
AVKAGSTEVDKVRNAGOSFKAPSCYTLVNDKSNHHLKPYMIGEIQDDQGEVVM
KTEPPIRAPNSPFIPEGNEKKPEHALKSN"
/length=7826
/genes="PA4859"
/length=7826
/genes="PA4859"
/codon_start=1
/transl_table=11
/product="probable permealase of ABC transporter"
/db_xref="GI:9951131"
/translation="MRTAATSLAQVRRLLVGLDQAFSGLSGSLTLLAALGLAI
TYGLGVINMAGHMLGATSYVQVLLQFRLAPWALYPLLLPFAVFCVSAIGM
ALERTVIRHGLRPLETLLATGIVSLVLIQVRLMFGAONVEANPAMLSGGIOVLN
FLVPMNRIVIGFALLVLLTLLNRTGLNRAVYVTRNNMAACCGVPTGRVMDLA
LGLSGAGLGVALSOGVNGPDLGGYIIDSFLVVLGVGLGAGSVLAALFGLVA
NKILEPQIGAVLGKILILALVLFIQKRPQGLFALKGRVID"
/length=8776
/genes="PA4860"
/length=8776
/genes="PA4860"
/codon_start=1
/transl_table=11
/product="probable permealase of ABC transporter"
/db_xref="GI:9951132"
/translation="MNQPLVTLARAGRVYATALGAAPATLLVMPLLHLLPADSAL
HLSATLITGLKICLAIYALADLVNMGAGLHGLFLFAGGAGMGLMQRASG
DSLPATFTLSWNLPTWAGTQHLFWALCLVLPVGLAFAGFAPFARIRGVYFS
IMQALTFAGMLLFNRTGFGNGNFTDFRSILGFITAPTRAALFIATVALLAAS
LLGLWRLARSKFRVLTALDAENRLMFCGYPKLVFTVLTSLVLCGLAGALFVPO
VGINPSEMSPTNISIAAVVAVVLCGRGTLLGLLGGVAGVWAKSWFTVAFPEYMLFPL
GALFIVVTLYLPRGLGLLRKEQ"
/length=9852
/genes="PA4861"
/length=9852
/genes="PA4861"
/codon_start=1
/transl_table=11
/product="probable ATP-binding component of ABC transporter"
/db_xref="GI:9951133"
/translation="MRTPHMLPEAPFHPGLDPTAGSGRDAIGLAATSGRDLARHGTI
LSUEDINVSDFGFKALDRLTYIGVGLERCIIGPNAGAKTLLMDVITGTRPQSGTAY
FGDTLDRMSEVQIAQIGRGRFQPTVFVLEALSIFVFNLELAQADKSWASLARLID
QTREREIVLATIRLLESQRPAGLLSHGQKQFLTEIGMLLVQEPOLLIDPEPVAGMT
DAETFTABLFSKLARKHSLMVEHDMGVGVSADHVTVLHQHVLAEGLAEVQADE
QVTEVYLSR"
/length=1768
BASE COUNT 1768 a 3916 c 3416 g 1707 t
ORIGIN

```

```

Query Match          0.9%; Score 22; DB 1; Length 10807;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 gttcttcgacctgtagctagct 87
|||||
Db 6704 gttcttcgacctgtagctagct 6683

RESULT 4
AC021564/c
LOCUS AC021564.3 GI:7417805
DEFINITION Homo sapiens clone RP11-772J6, WORKING DRAFT SEQUENCE, 17 unordered
pieces.
AC021564
VERSION AC021564.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 170970)
Homo sapiens, clone RP11-772J6
Unpublished
2 (bases 1 to 170970)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenescor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Ollivar,T.M., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severly,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Teafaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6899714.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5687
Center clone name: 772_J_6
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160384 bases at least Q40
Consensus quality: 165238 bases at least Q30
Consensus quality: 167127 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 169370; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1005: contig of 1005 bp in length
* 1006 1105: gap of 100 bp
* 1106 3028: contig of 1923 bp in length
* 3029 3128: gap of 100 bp
* 3129 4938: contig of 1810 bp in length
* 4939 5038: gap of 100 bp
* 5039 7205: contig of 2167 bp in length
* 7206 7305: gap of 100 bp
* 7306 10633: contig of 3328 bp in length
* 10634 10733: gap of 100 bp
* 10734 15053: contig of 4320 bp in length
* 15054 15153: gap of 100 bp
* 15154 18965: contig of 3812 bp in length
* 18966 19065: gap of 100 bp
* 19066 25452: contig of 6387 bp in length
* 25453 25552: gap of 100 bp
* 25553 33888: contig of 8336 bp in length
* 33889 33988: gap of 100 bp
* 33989 44820: contig of 10832 bp in length
* 44821 44920: gap of 100 bp
* 44921 53118: contig of 8198 bp in length
* 53119 53218: gap of 100 bp
* 53219 64484: contig of 11266 bp in length
* 64485 64584: gap of 100 bp
* 64585 80284: contig of 15700 bp in length
* 80285 80384: gap of 100 bp
* 80385 97652: contig of 17268 bp in length
* 97653 97752: gap of 100 bp
* 97753 116038: contig of 18266 bp in length
* 116039 116138: gap of 100 bp
* 116139 136813: contig of 20675 bp in length
* 136814 136913: gap of 100 bp
* 136914 170970: contig of 34057 bp in length.

FEATURES

source

1. .170970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ppl1-772J6"
/clone_lib="RPCI-11 Human Male BAC"
1. .1005
/note="assembly_fragment"
1106. .3028
/note="assembly_fragment"
3129. .4938
/note="assembly_fragment"
5039. .7205
/note="assembly_fragment"
7306. .10633
/note="assembly_fragment"
10734. .15053
/note="assembly_fragment"
15154. .18965
/note="assembly_fragment"
clone_end.T7
vector_side:left"
19066. .25452
/note="assembly_fragment"
25553. .33888
/note="assembly_fragment"
33989. .44820
/note="assembly_fragment"
44921. .53118
/note="assembly_fragment"
53219. .64484
/note="assembly_fragment"
64585. .80284

misc_feature
/note="assembly_fragment"
80385. .97652
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
97753. .116038
/note="assembly_fragment"
116139. .136813
/note="assembly_fragment"
136914. .170970
/note="assembly_fragment"
BASE COUNT 51652 a 32898 c 32559 g 52259 t 1602 others
ORIGIN

Query Match 0.9% Score 22; DB 2; Length 170970;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2225 ataacatgggatttaaaagga 2246
|||||
Db 116523 ATAACATGGGATTTAAAGGA 116502

RESULT 5
ARI41635/c
LOCUS ARI41635 30 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 4 from patent US 6146869.
ACCESSION ARI41635
VERSION ARI41635.1 GI:15101151
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Harris, P. and Brown, K.M.
TITLE Polypeptides having phospholipase B activity and nucleic acids
encoding same
JOURNAL Patent: US 6146869-A 4 14-NOV-2000;
FEATURES Location/Qualifiers
source
1. .30
/organism="unknown"
BASE COUNT 12 a 3 c 6 g 9 t
ORIGIN

Query Match 0.9% Score 21; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2030 ccatacccttttgggtataagt 2050
|||||
Db 30 CCATACCTTTTGGGTATAAGT 10

RESULT 6
AX127524/c
LOCUS AX127524 30 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 4 from Patent WO0129222.
ACCESSION AX127524
VERSION AX127524.1 GI:14134229
KEYWORDS Aspergillus oryzae.
SOURCE Aspergillus oryzae.
ORGANISM Aspergillus oryzae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 30)
AUTHORS Harris, P. and Brown, K.M.
TITLE Polypeptides having phospholipase B activity and nucleic acids
encoding same
JOURNAL Patent: WO 0129222-A 4 26-APR-2001;
NOVO NORDISK BIOTECH, INC. (US)
FEATURES Location/Qualifiers

```

source
1. .30
/organism="Aspergillus oryzae"
/db_xref="taxon:5062" 9 t

BASE COUNT      12 a      3 c      6 g      9 t
ORIGIN

Query Match      0.9%; Score 21; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2030 ccatacttttcggataagt 2050
|||||
Db 30 CCATACCTTCGGTATAAGT 10

RESULT 7
AC011538      38627 bp      DNA      PRI      21-JUN-2001
DEFINITION Homo sapiens chromosome 19 clone LLNLF-252C1, complete sequence.
ACCESSION AC011538
VERSION AC011538.6 GI:14518401
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 38627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38627)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 38627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On May 27, 2000 this sequence version replaced gi:7711529.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .38627
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLF-252C1"

BASE COUNT      9713 a      7023 c      7436 g      14455 t
ORIGIN

Query Match      0.9%; Score 21; DB 9; Length 38627;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2196 tattgtatttcattcaatgaa 2216
|||||
Db 9616 TATTGTATTTCATTCAATGAA 9636

RESULT 8
AC021090/c      40611 bp      DNA      PRI      27-MAY-2000
DEFINITION Homo sapiens chromosome 19 clone LLNLF-175G4, complete sequence.
ACCESSION AC021090
VERSION AC021090.5 GI:8099267

```

```

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40611)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40611)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 40611)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On May 27, 2000 this sequence version replaced gi:7711670.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .40611
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLF-175G4"

BASE COUNT      14376 a      7707 c      7409 g      11119 t
ORIGIN

Query Match      0.9%; Score 21; DB 9; Length 40611;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2196 tattgtatttcattcaatgaa 2216
|||||
Db 17399 TATTGTATTTCATTCAATGAA 17379

RESULT 9
AC092279      105538 bp      DNA      HTG      03-JUL-2001
LOCUS Homo sapiens chromosome 19 clone CTD-2017D11, WORKING DRAFT
DEFINITION SEQUENCE, 11 unordered pieces.
ACCESSION AC092279
VERSION AC092279.1 GI:14589468
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 105538)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 105538)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 634160, BC599184

```

```

Center clone name: CITB-HI_2017D11
-----
Summary Statistics
Consensus quality: 91971 bases at least Q40
Consensus quality: 100707 bases at least Q30
Consensus quality: 102669 bases at least Q20
Estimated insert size: 140000; agarose-fp estimation
Estimated insert size: 104538; sum-of-contigs estimation
Quality coverage: 5.05 in Q20 bases; agarose-fp estimation
Quality coverage: 6.76 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1048: contig of 1048 bp in length
* 1049: gap of unknown length
* 1149: contig of 1034 bp in length
* 2183: gap of unknown length
* 2283: contig of 2409 bp in length
* 4692: gap of unknown length
* 4792: contig of 2495 bp in length
* 7287: gap of unknown length
* 7386: contig of 4403 bp in length
* 7387: gap of unknown length
* 11790: gap of unknown length
* 11899: contig of 6745 bp in length
* 18634: gap of unknown length
* 18734: gap of unknown length
* 31948: contig of 13214 bp in length
* 32049: gap of unknown length
* 41163: contig of 9115 bp in length
* 41264: gap of unknown length
* 56571: contig of 15308 bp in length
* 56572: gap of unknown length
* 56672: contig of 18655 bp in length
* 75327: gap of unknown length
* 75426: gap of unknown length
* 75427: contig of 30112 bp in length.
*
* Location/Qualifiers
* 1..105538
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="19"
* /clone="CTD-2017D11"
* /clone_lib="CalTech human BAC library D"
* /clone_1ib="22832 g 30654 t 1024 others"
BASE COUNT 28572 a 22456 c 22832 g 30654 t 1024 others
ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 105538;
Best local similarity 100.0%; Pred. No. 8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2196 tatgtatattcaataatgaa 2216
|||||
Db 29427 TATGTATATTCATTAATGAA 29447

RESULT 10
AC091983/c AC091983 153173 bp DNA HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-559F22, WORKING DRAFT
DEFINITION SEQUENCE, 41 unordered pieces..
ACCESSION AC091983.1 GI:14333919
VERSION AC091983
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153173)
AUTHORS DOE Joint Genome Institute.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 153173)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 621163
Center clone name: RPCI-11_559F22
-----
Summary Statistics
Consensus quality: 106792 bases at least Q40
Consensus quality: 130951 bases at least Q30
Consensus quality: 136968 bases at least Q20
Estimated insert size: 196170; agarose-fp estimation
Estimated insert size: 149173; sum-of-contigs estimation
Quality coverage: 5.72 in Q20 bases; agarose-fp estimation
Quality coverage: 7.52 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1113: contig of 1113 bp in length
* 1114: gap of unknown length
* 1214: contig of 1098 bp in length
* 2312: gap of unknown length
* 2412: contig of 1589 bp in length
* 4001: gap of unknown length
* 4101: contig of 1297 bp in length
* 5398: gap of unknown length
* 5498: contig of 1094 bp in length
* 5592: gap of unknown length
* 6692: contig of 1185 bp in length
* 7877: gap of unknown length
* 7976: contig of 1418 bp in length
* 9395: gap of unknown length
* 9494: contig of 1471 bp in length
* 10965: gap of unknown length
* 11066: contig of 1976 bp in length
* 13042: gap of unknown length
* 13142: contig of 1235 bp in length
* 14377: gap of unknown length
* 14777: contig of 1212 bp in length
* 15689: gap of unknown length
* 15789: contig of 1996 bp in length
* 17785: gap of unknown length
* 17885: contig of 1843 bp in length
* 19727: gap of unknown length
* 19728: contig of 1028 bp in length
* 19828: gap of unknown length
* 20856: contig of 1165 bp in length
* 20956: gap of unknown length
* 22121: contig of 1228 bp in length
* 22221: gap of unknown length
* 23448: contig of 2127 bp in length
* 23549: gap of unknown length
* 25675: contig of 1321 bp in length
* 25676: gap of unknown length
* 25776: contig of 1321 bp in length
* 27097: gap of unknown length
* 27197: contig of 2142 bp in length
* 29338: gap of unknown length
* 29339: contig of 1500 bp in length
* 30939: gap of unknown length
* 31038: contig of 1624 bp in length
* 31039: gap of unknown length

```

```

* 32663 32762: gap of unknown length
* 32763 34189: contig of 1427 bp in length
* 34190 34289: gap of unknown length
* 34290 35589: contig of 1300 bp in length
* 35590 35689: gap of unknown length
* 35690 37304: contig of 1615 bp in length
* 37305 37404: gap of unknown length
* 37405 39017: contig of 1613 bp in length
* 39018 39117: gap of unknown length
* 39118 40664: contig of 1547 bp in length
* 40665 40764: gap of unknown length
* 40765 42207: contig of 1443 bp in length
* 42208 42307: gap of unknown length
* 42308 43822: contig of 1515 bp in length
* 43823 43922: gap of unknown length
* 43923 46593: contig of 2671 bp in length
* 46594 46693: gap of unknown length
* 46694 49219: contig of 2526 bp in length
* 49220 49319: gap of unknown length
* 49320 51644: contig of 2325 bp in length
* 51645 51744: gap of unknown length
* 51745 54629: contig of 2885 bp in length
* 54630 54729: gap of unknown length
* 54730 56080: contig of 1351 bp in length
* 56081 56180: gap of unknown length
* 56181 59449: contig of 3269 bp in length
* 59450 59549: gap of unknown length
* 59550 64245: contig of 4696 bp in length
* 64246 64345: gap of unknown length
* 64346 69497: contig of 5152 bp in length
* 69498 69597: gap of unknown length
* 69598 72251: contig of 2654 bp in length
* 72252 72351: gap of unknown length
* 72352 77749: contig of 5398 bp in length
* 77750 77849: gap of unknown length
* 77850 87328: contig of 9479 bp in length
* 87329 87428: gap of unknown length
* 87429 103483: contig of 16055 bp in length
* 103484 103583: gap of unknown length
* 103584 153173: contig of 49590 bp in length.
FEATURES
    source
        location/Qualifiers
            1. 153173
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone_lib="RPCI human BAC library 11"
BASE COUNT 45483 a 28901 c 29032 g 45744 t 4013 others
ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 153173;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2196 tattgtatattcattatga 2216
|||||
Db 91857 TATTGTATTTCATTATGA 91837

RESULT 11
AC072046/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-450H5, WORKING DRAFT SEQUENCE, 11-NOV-2000
ACCESSION AC072046
VERSION AC072046.6 GI:11128191
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184606)

```

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Brieva,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dunn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsil,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,K., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 184606)

Worley,K.C.

Direct Submission

Submitted (07-JUN-2000)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2000 this sequence version replaced gi:9625287.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBPB
Center clone name: RP11-450H5
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 83% of reads
Chemistry: Dye-terminator Big Dye; 83% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 163304 bases at least Q40
Consensus quality: 174355 bases at least Q30
Consensus quality: 178508 bases at least Q20
Estimated insert size: 178126; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1	21440:	contig of 21440 bp in length
21441	21540:	gap of unknown length
21541	40789:	contig of 19249 bp in length
40790	40889:	gap of unknown length
40890	57958:	contig of 17109 bp in length
57959	58098:	gap of unknown length
58099	72182:	contig of 14084 bp in length
72183	72282:	gap of unknown length
72283	81930:	contig of 9648 bp in length
81931	82030:	gap of unknown length
82031	90915:	contig of 8885 bp in length
90916	91015:	gap of unknown length
91016	95604:	contig of 4589 bp in length
95605	95704:	gap of unknown length
95705	104642:	contig of 8938 bp in length
104643	104742:	gap of unknown length
104743	113644:	contig of 8902 bp in length
113645	113744:	gap of unknown length
113745	120572:	contig of 6828 bp in length
120573	120672:	gap of unknown length
120673	127681:	contig of 7009 bp in length
127682	127781:	gap of unknown length
127782	13381:	contig of 6070 bp in length
13382	133951:	contig of unknown length
133952	139649:	contig of 3698 bp in length
139650	139749:	gap of unknown length
139750	145151:	contig of 5402 bp in length
145152	145251:	gap of unknown length
145252	150958:	contig of 5707 bp in length
150959	151058:	gap of unknown length
151059	15261:	contig of 4203 bp in length
15262	15361:	gap of unknown length
15362	159825:	contig of 4464 bp in length
159826	159925:	gap of unknown length
159926	164213:	contig of 4288 bp in length
164214	164313:	gap of unknown length
164314	167956:	contig of 3643 bp in length
167957	168056:	gap of unknown length
168057	171292:	contig of 3236 bp in length
171293	171392:	gap of unknown length
171393	174510:	contig of 3118 bp in length
174511	174610:	gap of unknown length
174611	176567:	contig of 1957 bp in length
176568	176667:	gap of unknown length
176668	178426:	contig of 1759 bp in length
178427	178526:	gap of unknown length
178527	180699:	contig of 2173 bp in length
180700	180799:	gap of unknown length
180800	182025:	contig of 1226 bp in length
182026	182125:	gap of unknown length
182126	183460:	contig of 1335 bp in length
183461	183560:	gap of unknown length
183561	184606:	contig of 1046 bp in length.

FEATURES
Source
1. 184606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-450H5"
/chromosome="3"
60191 a 32101 c 30675 g 58997 t 2642 others

Query Match 0.9% Score 21; DB 2; Length 184606;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 712 gatattcatgtggaataat 732

Db 27702 GATATTGATGTGGAATAAT 27682

RESULT 12
LOCUS AC067865 200689 bp DNA HTG 25-JUN-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-314N9 map 2, WORKING DRAFT
SEQUENCE, 15 unordered pieces.

AC067865
AC067865.2 GI:87051161
HTG: HTGS_PHASE1; HTGS_DRAFT.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 200689)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

JOURNAL
Homo sapiens chromosome 2, clone RP11-314N9
Unpublished

REFERENCE
2 (bases 1 to 200689)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boucknight, B., Brown, A., Burkett, G.,
Campopiano, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, P., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 25, 2000 this sequence version replaced gi:7651917.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L9065

Center clone name: 314_N_9
Summary Statistics
Sequencing vector: M13: M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 189802 bases at least Q40
Consensus quality: 195895 bases at least Q30

Consensus quality: 198051 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 199289; sum-of-contigs

Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 15 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1975: contig of 1975 bp in length
 * 1976 2075: gap of 100 bp
 * 2076 4372: contig of 2297 bp in length
 * 4373 4472: gap of 100 bp
 * 4473 8606: contig of 4134 bp in length
 * 8607 8708: gap of 100 bp
 * 8707 16318: contig of 7612 bp in length
 * 16319 16418: gap of 100 bp
 * 16419 22960: contig of 6542 bp in length
 * 22961 23060: gap of 100 bp
 * 23061 32721: contig of 9661 bp in length
 * 32722 32821: gap of 100 bp
 * 32822 40636: contig of 7815 bp in length
 * 40637 40736: gap of 100 bp
 * 40737 54297: contig of 13561 bp in length
 * 54298 54397: gap of 100 bp
 * 54398 70328: contig of 15931 bp in length
 * 70329 70428: gap of 100 bp
 * 70429 87196: contig of 16768 bp in length
 * 87197 87296: gap of 100 bp
 * 87297 104624: contig of 17328 bp in length
 * 104625 104724: gap of 100 bp
 * 104725 125960: contig of 21236 bp in length
 * 125961 126060: gap of 100 bp
 * 126061 146883: contig of 20823 bp in length
 * 146884 146983: gap of 100 bp
 * 146984 174757: contig of 27774 bp in length
 * 174758 174857: gap of 100 bp
 * 174858 200689: contig of 25832 bp in length.

.FEATURES

source
 1. .200689
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone_lib="RP11-11 Human Male BAC"
 1. .1975
 /note="assembly_fragment"
 2076. .4372
 /note="assembly_fragment"
 4473. .8606
 /note="assembly_fragment"
 8707. .16318
 /note="assembly_fragment"
 16419. .22960
 /note="assembly_fragment"
 23061. .32721
 /note="assembly_fragment"
 32822. .40636
 /note="assembly_fragment"
 40737. .54297
 /note="assembly_fragment"
 54398. .70328
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 70429. .87196
 /note="assembly_fragment"
 87297. .104624
 /note="assembly_fragment"
 104725. .125960
 /note="assembly_fragment"
 126061. .146883
 /note="assembly_fragment"
 146984. .174757
 /note="assembly_fragment"
 174858. .200689
 /note="assembly_fragment"

BASE COUNT 68219 a 34336 c 34359 g 62374 t 1401 others
 ORIGIN
 Query Match 0.9%; Score 21; DB 2; Length 200689;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 712 gatattcatgtggaataat 732
 |||||
 Db 199512 GATATTGATGTGGAATAAT 199532
 |||||
 RESULT 13
 AC067900 201312 bp DNA HTG 09-MAY-2001
 LOCUS Homo sapiens chromosome 8 clone RP11-775M3 map 8, WORKING DRAFT
 DEFINITION SEQUENCE, 5 unordered pieces.
 AC067900
 VERSION AC067900.3 GI:12545341
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 201312)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 8, clone RP11-775M3
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 201312)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
 Boquslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campolunghi,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 TITLE
 JOURNAL
 COMMENT Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 26, 2001 this sequence version replaced gi:8138718.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7524
 Center clone name: 775_M_3
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 41% of reads
 Sequencing vector: Plasmid; n/a; 59% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 198330 bases at least Q40
 Consensus quality: 199387 bases at least Q30

Consensus quality: 200109 bases at least Q20
 Insert size: 210000; agarose-fp
 Insert size: 200912; sum-of-contigs
 Quality coverage: 9.1 in Q20 bases; agarose-fp
 Quality coverage: 9.5 in Q20 ba.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 77871: contig of 77871 bp in length
 * 77872 77971: gap of 100 bp
 * 77972 83159: contig of 5188 bp in length
 * 83160 83259: gap of 100 bp
 * 83260 115460: contig of 32201 bp in length
 * 115461 115560: gap of 100 bp
 * 115561 158995: contig of 43435 bp in length
 * 158996 159095: gap of 100 bp
 * 159096 201312: contig of 42217 bp in length.

FEATURES

Source

1..201312 Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-775M3"

1..77871 Human Male BAC"

misc_feature

1..77871
 /note="assembly_fragment"

clone_end:SF6
 vector_side:left"

misc_feature

77972..83159
 /note="assembly_fragment"

misc_feature

83260..115460
 /note="assembly_fragment"

misc_feature

115561..158995
 /note="assembly_fragment"

misc_feature

159096..201312
 /note="assembly_fragment"

BASE COUNT 65566 a 34813 c 36561 g 63964 t 408 others

ORIGIN

Query Match 0.98; Score 21; DB 2; Length 201312;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2222 ttgataacatgggtattaa 2242

|||||

Db 70363 TTGATACATGGGATTAA 70383

RESULT 14

AC073534

LOCUS AC073534 209910 bp DNA HTG 20-APR-2001

DEFINITION Homo sapiens chromosome 19 clone CTD-2027119, WORKING DRAFT

SEQUENCE, 74 unordered pieces.

AC073534

AC073534.3 GI:13699747

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE

human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 209910)

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 19

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 209910)

AUTHORS DOE Joint Genome Institute.

TITLE
JOURNAL

COMMENT

Direct Submission
 Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Apr 20, 2001 this sequence version replaced gi:9554695.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 638128, BC603152
 Center clone name: CTFB-HI_2027119

Summary Statistics

Consensus quality: 121533 bases at least Q40
 Consensus quality: 142219 bases at least Q30
 Consensus quality: 152387 bases at least Q20
 Estimated insert size: 141340; agarose-fp estimation.
 Estimated insert size: 202610; sum-of-contigs estimation
 Quality coverage: 9.39 in Q20 bases; agarose-fp estimation
 Quality coverage: 6.55 in Q20 bases; sum-of-contigs estimation.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1303: contig of 1303 bp in length
 * 1304 1403: gap of unknown length
 * 1404 3247: contig of 1844 bp in length
 * 3248 3347: gap of unknown length
 * 3348 4388: contig of 1021 bp in length
 * 4389 4468: gap of unknown length
 * 4469 5629: contig of 1161 bp in length
 * 5630 5729: gap of unknown length
 * 5730 6752: contig of 1023 bp in length
 * 6753 6852: gap of unknown length
 * 6853 8296: contig of 1444 bp in length
 * 8297 8396: gap of unknown length
 * 8397 9437: contig of 1041 bp in length
 * 9438 9537: gap of unknown length
 * 9538 10687: contig of 1149 bp in length
 * 10688 10786: gap of unknown length
 * 10787 11792: contig of 1006 bp in length
 * 11793 11892: gap of unknown length
 * 11893 12911: contig of 1019 bp in length
 * 12912 13011: gap of unknown length
 * 13012 14060: contig of 1049 bp in length
 * 14061 14160: gap of unknown length
 * 14161 15213: contig of 1053 bp in length
 * 15214 15313: gap of unknown length
 * 15314 16503: contig of 1190 bp in length
 * 16504 16603: gap of unknown length
 * 16604 17729: contig of 1126 bp in length
 * 17730 17829: gap of unknown length
 * 17830 18997: contig of 1168 bp in length
 * 18998 19097: gap of unknown length
 * 19098 20118: contig of 1021 bp in length
 * 20119 20218: gap of unknown length
 * 20219 21253: contig of 1035 bp in length
 * 21254 21353: gap of unknown length
 * 21354 22435: contig of 1082 bp in length
 * 22436 22535: gap of unknown length
 * 22536 23539: contig of 1004 bp in length
 * 23540 23639: gap of unknown length
 * 23640 24791: contig of 1152 bp in length
 * 24792 24891: gap of unknown length
 * 24892 26023: contig of 1132 bp in length
 * 26024 26123: gap of unknown length
 * 26124 27261: contig of 1138 bp in length
 * 27262 27361: gap of unknown length

under conditions of nitrogen deprivation, which is the normal situation for *B. cinerea* during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

FEATURES
source Location/Qualifiers
1..240
/organism="Botryotinia fuckeliana"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W13D061"
BASE COUNT 74 a 50 c 65 g 51 t
ORIGIN

Query Match 0.9%; Score 20; DB 8; Length 240;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 648 ctggaaggacaaaggtcgaga 667
|||||
Db 102 CTTGAAGGACAAAGGTCGAGA 121

Search completed: December 4, 2001, 18:48:10
Job time: 11394 sec

